LEABHARLANN CHOLÁISTE NA TRÍONÓIDE, BAILE ÁTHA CLIATH | TRINITY COLLEGE LIBRARY DUBLIN
Ollscoil Átha Cliath | The University of Dublin

Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Regulation of Cancer Stem Cell Differentiation by Genes and MicroRNAs

Sebastian Vencken, B.Sc. (Hon)

Department of Histopathology and Morbid Anatomy
School of Medicine
University of Dublin, Trinity College

A thesis submitted to University of Dublin, Trinity College for the degree of Doctor of Philosophy

September 2012

Under the supervision of Professor John O'Leary and Dr. Michael Gallagher

TRINITY COLLEGE
2 4 MAY 2013
LIBRARY DUBLIN
Thesis 10023

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

Sebastian Vencken

Summary

Since the discovery of a stem cell phenotype in cancer, specific tumour cells with this phenotype, often called cancer stem cells (CSCs), are now widely accepted as the progenitors of oncogenesis, proliferation, treatment resistance and metastasis. For this reason a lot of research is focused on the characterisation of these cells and their exact roles in the stages of cancer. Most of today's therapies focus on eliminating the tumour bulk, while frequently missing a small subpopulation of CSCs, which engender the recurrence and further propagation of the disease. A growing consensus supports CSC-targeted therapy as a viable method of fully eradicating cancer, while also eliminating the frequent incidence of recurrence, dissemination and treatment-resistance of the disease. The elimination of CSCs is very challenging because of the extensive abilities of these cells to avoid detection and treatment. The deliberate induction of CSC differentiation is investigated as a possible therapeutic mechanism, as it can decrease the proliferation of CSCs and increase their sensitivity to treatment. CSC differentiation introduces a promising method to control cancer and is the main focus of this thesis.

MicroRNAs are short non-protein-coding oligonucleotides of approximately 21-22 nucleotides in length. Their dominant function is the selective post-transcriptional repression of genes by binding to the 3' untranslated region of their mRNAs and subsequently preventing protein translation. Approximately 1500 microRNA sequences have been detected in the human genome, which have been estimated to regulate more than 60% of all protein-coding genes. This microRNA network regulates crucial biological processes and their aberrant expression has frequently been attributed with potent oncogenic or tumour repressive characteristics.

In this study, two human embryonal carcinoma cell lines were used as CSC models; the nullipotent 2102Ep and pluripotent NTera-2 cell lines. Whereas NTera-2 cells rapidly differentiate upon the introduction of morphogens, 2102Ep cells have an innate resistance to terminal differentiation. Both cell lines characteristically express a large number of human embryonic stem cell (hESC) markers, among which the three core pluripotency markers, OCT4, SOX2 and NANOG. Both cell lines partially resemble the hESC phenotype and have previously been suggested as alternative hESC models.

The aims of this study were to find genes and microRNAs that had crucial and preferably specific roles in the differentiation programs of 2102Ep and NTera-2 cells. Laboratory protocols were designed and optimised for the transient silencing of genes and transient overexpression of microRNAs. For functional analyses, six genes (SOX2, DPPA3, SHISA2, RBM47, AK3L1 and VSNL1) and five microRNAs (miR-10a, -99a, -184, -219 and miR-335) were selected with the criteria that

they were highly deregulated upon early differentiation of NTera-2 cells and potentially had important functions pertaining to CSC or embryonal carcinoma differentiation.

The induced silencing of SOX2 resulted in the differentiation of both 2102Ep and NTera-2 cell lines, while the induced upregulation of miR-335 resulted in the differentiation of the latter cell line. This was measured by a marked reduction of mRNA expression of the three core pluripotency markers, OCT4, SOX2 and NANOG, and, for SOX2 silencing experiments, the reduction of cell surface alkaline phosphatase activity. The expression of some early differentiation markers was measured in 2102Ep cells and these were found to upregulated.

SOX2 and OCT4 were also computationally predicted targets of miR-335. The protein expression of these genes was measured in miR-335-overexpressing NTera-2 cells. This showed a loss of SOX2 and OCT4 protein. When In parallel to protein expression, the transcript expression of OCT4 was downregulated compared to negative controls. Conversely, the transcript expression of SOX2 was upregulated in miR-335 overexpressing cells. Particularly for the latter case, this finding suggested that SOX2 was post-transcriptionally regulated, possibly by miR-335.

The SOX2-silenced 2101Ep and NTera-2 cells were profiled for differential whole-genome and whole-microRNAome expression to further elucidate the role of the SOX2 regulatory network in CSC differentiation. This revealed that in both these cell lines, SOX2 regulated a substantial number of genes related to the Wnt and TGFβ signalling pathways, two crucial pathways to CSC biology. A large proportion of microRNAs were also deregulated by SOX2-silencing, among which the most notable were the downregulation of members of the oncogenic miR-371-373 cluster in NTera-2 cells and the C19MC polycistron in 2102Ep cells. Besides these findings, several microRNAs that were validated, direct inhibitors of SOX2 in other cell lines, were differentially expressed, suggesting the existence of novel SOX2:miRNA autoregulatory loops.

The findings in this study have highlighted novel members of the embryonal carcinoma differentiation network and potential therapeutic targets in this cancer and perhaps CSCs in general. This study has also further detailed the differences between the 2102Ep and NTera-2 cell lines and tested their resemblance to hESCs.

Table of Contents

Acknowledgements I

List of Tables V

List of Figures IX

Abbreviations XV

Publications XIX

1 — Introduction 2

1.1 — The Stem Cell Theory of Cancer 2

1.2 — The Role of Cancer Stem Cells in Carcinogenicity 6

1.3 — Cancer Stem Cells and Metastasis 8

1.4 — Cancer Stem Cell-targeted Cancer Therapy and Its Implications 11

1.5 — The Origins of Cancer Stem Cells and the Heterogeneous versus Stochastic Model of Cancer 16

1.6 — MicroRNA 18

1.7 — Biosynthesis and molecular processing of MicroRNA 19

1.8 — Post-Transcriptional Regulation of Protein Synthesis by MiRNA 20

1.8.1 — The RNA-induced Silencing Complex 20

1.8.2 — Targeting mRNAs 21

1.9 — Regulation of MicroRNA 23

1.10 — Additional Functions of miRNA 25

1.11 — MicroRNA and Cancer 26

1.12 — MiRNA regulation in Cancer Stem Cells and Embryonal Carcinoma 30

1.13 — The Cancer Stem Cell Model of this Study 33

1.14 — Hypotheses 34

2 — Materials and Methods 38

2.1 — Cell Culture 38

2.1.3 — Cell Lines 38

2.1.4 — Initial Cell Seeding 38

2.1.5 — Cell Culture Maintenance 39

2.1.6 — Long-Term Cell Storage 40

2.1.7 — Cell Counting 40

2.1.8 — Cell Viability Assay 40

2.1.9 — Alkaline Phosphatase Activity Assay 41

2.1.10 — Mycoplasma Contamination Testing 42

2.2 — Transfection of Nucleic Acids 43

2.2.1 — Methodology 43

2.2.1.1 — Nucleic Acid Constructs 43

2.2.1.2 — Transfection Agents 44

2.2.1.3 — Transfection Methods 44

2.2.1.4 — Transfection Conditions 45

2.2.2 — Methods 46

2.2.2.1 — Cell Media 46

2.2.2.2 — Reverse Transfection Optimisation on 96-well Cell Culture Plates 47

2.2.2.3 — Forward Transfection Optimisation on 96-well Cell Culture Plates 48

2.2.2.4 — Forward Transfections on 96-well Cell Culture Plates 50

2.2.2.5 — Forward Transfections on 24-well Cell Culture Plates 50

2.2.2.6 — Forward Transfections on 6-well Cell Culture Plates 51

2.3 — Total RNA Isolation and Purification 52

2.3.1 — Methodology 52

2.3.1.1 — RNA Isolation with TaqMan® Cells-to-Ct™ Gene Expression Kit 52

2.3.1.2 — TaqMan® Cells-to-Ct ™ Control Kit 52

2.3.1.3 — Kits for RNA Purification 53

2.3.2 — Methods 54
2.3.2.1 — TaqMan® Cells-to-Ct Gene Expression Kit 54
2.3.2.2 — miRVana™ miRNA Isolation kit 54
2.3.2.3 — miRVana™ PARIS kit 55
2.3.2.4 — MagMax™-96 Total RNA Isolation kit 56
2.3.2.5 — DNA Removal 57
2.4 — RNA Quantification and Qualification 58
2.4.1 — Methodology 58
2.4.1.1 — NanoDrop 2000 58
2.4.1.2 — 2100 Bioanalyzer 58
2.4.2 — Methods 59
2.4.2.1 — NanoDrop 2000 59
2.4.2.2 — 2100 Bioanalyzer 59
2.5 — RNA Expression Analysis 60
2.5.1 — Methodology 60
2.5.1.1 — TaqMan® Quantitative Real-Time PCR 60
2.5.1.2 — TaqMan® Array Human microRNA 64
2.5.1.3 — Affymetrix Gene Array 65
2.5.2 — Methods 66
2.5.2.1 — TaqMan® Quantitative Real-Time PCR of mRNA 66
2.5.2.2 — Cells-to-Ct® Gene Expression 68
2.5.2.3 — TaqMan® Quantitative Real-Time PCR of miRNA 69
2.5.2.4 — TaqMan® Array Human MicroRNA 70
2.5.2.5 — Affymetrix GeneChip® Human Gene 1.0 ST Arrays 71
2.6 — Protein Expression Analysis 75
2.6.1 — Methodology 75
2.6.1.1 — Western Blot 75
2.6.2 — Methods 76

2.6.2.1 — Protein Extraction 76
2.6.2.2 — Protein Quantification 76
2.6.2.3 — SDS-PAGE 77
2.6.2.4 — Protein Transfer to Membranes 78
2.6.2.5 — Membrane Blocking and Antibody Binding 79
2.6.2.6 — Antibody Inhibition 80
2.6.2.7 — Enhanced Chemiluminscence 80
2.6.2.8 — Membrane Stripping 80
2.7 — Statistical Analysis 81
2.7.1 — Computational Statistical Analysis 81
3 — Induced Upregulation of miRNA in 2102Ep and NTera-2 Cells 84
3.1 — Introduction 84
3.2 — Chapter Aims 88
3.3 — Materials and Methods 89
3.3.1 — Cell Culture 89
3.3.2 — Cells-to-Ct™ Gene Expression 89
3.3.2.1 — Cells-to-Ct™ Control Kit 89
3.3.3 — Pre-miR™ Starter Kit 91
3.3.3.1 — Baseline PTK9 Expression Analysis 91
3.3.4 — Reverse Transfection Optimisation 91
3.3.4.1 — Optimisation of Experimental Media 92
3.3.4.2 — Optimisation of Reverse Transfections 92
3.3.5 — Forward Transfection Optimisation 92
3.3.5.1 — Optimisation of Experimental Media 92
3.3.5.2 — Optimisation of Forward Transfections 93
3.3.5.3 — Cell Viability 93
3.3.6 — Functional Pre-miR™ Transfections 93
3.3.7 — Analysis of miRNA and Gene Expression 95

3.3.8 — Protein Expression Analysis 96
3.3.9 — Computational miRNA Target Prediction 96
3.4 — Results 98
3.4.1 — Pre-optimisation Control Experiments 98
3.4.1.1 — Optimisation of the Cells-to-Ct™ Gene Expression Kit 98
3.4.1.2 — Determining Baseline PTK9 Expression in Both Cell Lines 101
3.4.2 — Reverse Transfection Optimisation of Pre-miR™ miRNA Precursors 103
3.4.2.1 — Optimisation of Experimental Media 103
3.4.2.2 — Optimisation of Reverse Transfection in 2102Ep Cells 105
3.4.2.3 - Optimisation of Reverse Transfection in NTera-2 Cells 107
3.4.3 — Forward Transfection Optimisation of Pre-miR™ miRNA Precursors 109
3.4.3.1 — Optimisation of Experimental Media 109
3.4.3.2 — Optimisation of Forward Transfections 111
3.4.3.3 — Cytotoxic Effects of Pre-miR™ Transfection 114
3.4.4 — Functional Upregulation of miRNAs 116
3.4.5 — Induced Overexpression of Five miRNAs in 2102Ep cells 120
3.4.5.1 – Phenotypic Analysis 120
3.4.5.2 — Expression of Functionally Upregulation miRNA 124
3.4.5.3 — Expression Analysis of Pluripotency Markers 127
3.4.5.4 — Expression Analysis of Differentiation Markers 130
3.4.6 — Induced Overexpression of Five miRNAs in NTera-2 Cells 133
3.4.6.1 - Phenotypic Analysis 133
3.4.6.2 - Expression Analysis of Functionally Upregulated miRNA 136
3.4.6.3 - Expression Analysis of Pluripotency Markers 139
3.4.6.4 — Expression Analysis of Differentiation Markers 142
3.4.6.5 — Analysis of 6-day MiRNA Transfections in NTera-2 Cells 144
3.4.6.6 — Protein Expression of Genes Potentially Targeted by miR-335 147
3.5 — Discussion 149

3.5.1 — Successful Establishment of a Laboratory Protocol for Transient Ectopic MiRNA Overexpression in 2102Ep and NTera-2 Cell Lines. 149

3.5.2 — Functional Overexpression of MiR-335 and MiR-184 Induced Differentiation In ECCs 149

3.5.3 — OCT4 and SOX2 as Possible Targets of MiR-335 152

4 — Induced Silencing of Genes in 2102Ep and NTera-2 Cells 160

4.1 — Introduction 160

4.2 — Chapter Aims 163

4.3 — Materials and Methods 164

4.3.1 — Cell Culture 164

4.3.2 — Optimisation of RNAi 164

4.3.2.1 — Transfection of siRNA 164

4.3.2.2 — Gene Expression Analysis 166

4.3.2.3 — Protein Expression Analysis 166

4.3.3 — Functional Gene Silencing 168

4.3.3.1 — Functional Transfection of siRNA 168

4.3.3.2 - Gene Expression Analysis 169

4.3.3.3 — Protein Expression Analysis 170

4.4 — Results 171

4.4.1 — Optimisation of RNAi 171

4.4.1.1 — Gene Expression Analysis 171

4.4.1.2 — Protein Expression Analysis 174

4.4.1.3 — Cytotoxic Effects of siRNA Transfection 175

4.4.2 — Functional Silencing of SOX2 in 2102Ep Cells 177

4.4.2.1 — Phenotypic Analysis 177

4.4.2.2 — Gene Expression Analysis 178

4.4.2.3 — Assessing Differentiation with an Alkaline Phosphatase Assay 181

4.4.3 — Functional Silencing of SOX2 in NTera-2 Cells 183

4.4.3.1 — Phenotypic Analysis 183

4.4.3.2 — Gene Expression Analysis 185

4.4.3.3 — Assessing Differentiation with an Alkaline Phosphatase Assay 187
4.4.4 — Functional Silencing of Five siRNA in 2102Ep 188
4.4.4.1 — Phenotypic Analysis 190
4.4.4.2 — Gene Expression Analysis 192
4.4.4.3 — Protein Expression Analysis 195
4.4.4.4 — Expression Analysis of Pluripotency Markers 197
4.4.4.5 — Alkaline Phosphatase Assay 198
4.4.4.6 — Assessing Cell Viability 200
4.4.5 — Functional Silencing of Five siRNA in NTera-2 Cells 202
4.4.5.1 — Phenotypic Analysis 202
4.4.5.2 — Gene Expression Analysis 205
4.4.5.3 — Protein Expression Analysis 209
4.4.5.4 — Expression Analysis of Pluripotency Markers 210
4.4.5.5 — Alkaline Phosphatase Assay 212
4.4.5.6 — Assessing Cell Viability 214
4.5 — Discussion 215
4.5.1 — Successful Establishment of a Laboratory Protocol for Transient Gene Silencing in 2102Ep and NTera-2 Cell Lines 216
4.5.2 — Experimental Silencing of SOX2 Induces Differentiation in 2102Ep and NTera-2 Cell Lines 216
4.5.3 — Silencing of Five Selected Genes Did Not Differentiate 2102Ep and NTera-2 Cells 217
5 — Global Gene and MicroRNA Expression Analysis Embryonal Carcinoma 222
5.1 — Introduction 222
5.2 - Chapter Aims 224
5.3 — Materials and Methods 225
5.3.1 — SOX2 silencing in NTera-2 and 2102Ep Cells 225
5.3.1.1 — RNA Purification and Protein Extraction 225
5.3.1.2 — Assessment of RNA and Protein Quantity and Quality 225
5.3.1.3 — Gene Expression Analysis 226

5.3.1.4 — Protein Expression Analysis 227

5.3.2 — Whole Gene Expression Analysis with Gene Arrays 228

5.3.2.1 — Analysis of RNA Integrity 228

5.3.2.2 — Sample Preparation, Hybridisation and Scanning 228

5.3.2.3 — Quality Control 228

5.3.2.4 — Probe Signal Summarisation, Background Correction and Normalisation. 231

5.3.2.5 — Computational Analsysis of Differential Gene Expression 232

5.3.2.6 – Computational Analysis of Gene Ontology 232

5.3.3 - TaqMan® Array Human MicroRNA 235

5.3.3.1 — Optimisation of the Preamplification Stage 235

5.3.3.2 — Whole miRNA Expression Analysis 235

5.3.3.3 — Computational Analysis of Differential miRNA Expression 236

5.3.3.4 — Computational miRNA Target Prediction 237

5.4 — Results 238

5.4.4 — Silencing of SOX2 in 2102Ep and NTera-2 Cells 238

5.4.4.1 — SOX2 Expression in 2102Ep 239

5.4.4.2 — SOX2 Expression in NTera-2 Cells 240

5.4.5 - RNA Quality and Quantity Assessment 241

5.4.6 — Whole-Transcriptome Differential Expression Analysis with Gene Arrays 244

5.4.6.1 — Quality Control 244

5.4.6.2 — Analysis of Differential Gene Expression 250

5.4.6.3 — Validation of Gene Array Results by qRT-PCR 253

5.4.6.4 — Further Inspection of Differential Gene Analysis 256

5.4.6.5 — Cross-reference with External Studies 261

5.4.6.6 — Pathway Analysis 264

5.4.6.7 — Functional Gene and Annotation Analysis 266

5.4.7 - Whole miRNA Expression Analysis with TaqMan® Arrays 271

5.4.7.1 — Optimisation of the miRNA Preamplification Stage 271

5.4.7.2 — Quality Control 274

5.4.7.3 — Analysis of Differential miRNA Expression 278

5.4.7.4 — Computational Target Prediction for Differentially Regulated miRNA 283

5.4.7.5 — Cross-reference Analysis of mRNA and microRNA Expression Data 288

5.5 — Discussion 295

5.5.8 — Whole-genome Differential Expression 296

5.5.8.1 — Core Pluripotency Markers 296

5.5.8.2 — A Comparison of 2102Ep and NTera-2 Cell Differentiation by SOX2 Silencing . 296

5.5.8.3 — Directly SOX2-linked Genes 299

5.5.9 — Differentially Expressed miRNAs in 2102Ep and NTera-2 Cells 303

5.5.9.1 — MicroRNAs in Cancer Stem Cells 305

5.5.9.2 — MicroRNAs in Cancer and Embryonic Stem Cell Function 306

5.5.9.3 — The Identification of MicroRNA Targets by Computational Analysis 312

5.5.9.4 — Functional Relevance of MicroRNA Clusters 315

6 — General Discussion 321

6.1 — Discussion of the Findings 323

6.2 — Prospects for Future Work 332

References 338

Acknowledgments

Firstly, for his unwaivering support and encouragment, I would like to thank my supervisor Professor John O'Leary. His infectious enthusiasm has always been greatly uplifting, especially during the occasional dip or panick during my PhD. I would also like to sincerely thank my co-supervisor, Dr. Michael Gallagher. His support, insight and his personal focus on the best interests for his students and their work were invaluable to the completion of this project.

The past three-and-a-half years have been a professional, but also personal experience and the advice and guidance you both have given to me, I will always be greatful for and will not be forgotten.

I would like to sincerely thank Dr. Cara Martin for all her help and advice and for her invaluable assistance in making the completion of this PhD possible.

My thanks go to Dr. Sharon O'Toole and Professor Orla Sheils, for their expert advice and for bringing clarity when this was needed.

I would like to personally thank my other colleagues at Pathology Research Lab, Cathy, Helen, Aoife C, Brendan, Christine, Itunu, Loretto, Lynda, but also those who have moved to new pastures, Louise, Jamie, Salah, Yvonne, Adele, Gráinne, Rachael, Aoife F, Catriona and Britta. On so many occasions have you provided me with essential help and assistance in the laboratory, and I will always be grateful for this. The Coombe has been a delight to work in and I loved the craic and banter we had in and outside the lab.

I would also like to thank the people at the labs in James' Hospital, who have selflessly helped me with my work. I would like to thank Paul Smyth for his skillful advice and support in the laboratory. My thanks go to Gordon Blackshields for his excellent and much needed assisentance in the data analysis and statistics of this project. For providing and assisting me with the use of his laboratory, I would also like to extend my thanks Stephen Smith.

I thank my friends who have been around for me since I first stepped on Irish shores. You have made my life here as a undergrad and PhD student unforgettable and I regret having seen some of you leave. In addition, Paul, Leonard, Conor and Brian, sorry for having to put you through a grueling proof read, but you guys did a great job.

To my Dutch, Irish and French families, I truely thank you for all you love, kindness, empathy and support. You have made me the person I am and you have always been there for me, whether in Ireland, Holland or France.

Finally, to Caoimhe. This PhD would not have happened without you. Your enduring strength, love and committment has made me a better person and has given me the strength to get where I am now. You have always been there for me, during the good and bad times. This PhD has been quite an experience for both of us, but I'm glad to have completed it with you by my side.

List of Tables

Table 1.1	Cancer stem cell model versus stochastic cancer model.	17
Table 1.2	Structural characteristics miRNA-base therapeutics and delivery methods.	30
Table 1.3	The two CSC models used in this study.	33
Table 2.1	The advantages and disadvantages of different transfection methods.	45
Table 2.2	Different volumes of media used during the transfection experiments.	46
Table 2.3	Different transfection agents used at different volumes on 96-well plates.	47
Table 2.4	Different incubation times for transfection agents.	48
Table 2.5	Different volumes of Lipofectamine™ 2000 transfection agent for 2102Ep on 96-wells.	49
Table 2.6	RNA purification kits used in this study.	53
Table 2.7	Components of a 12% resolving gel used for SDS-PAGE.	77
Table 3.1	2102Ep and NTera-2 serial cell dilutions for the Cells-to-Ct™ Control Kit.	89
Table 3.2	GeneAmp® 9600 thermocycler program for reverse transcription of sample RNA with the Cells-to-Ct™ Gene Expression kit.	90
Table 3.3	GeneAmp® 9600 thermocycler program for qRT-PCR with the Cells-to-Ct™ Gene Expression kit.	91
Table 3.4	Product type and sequence of Pre-miR™ precursors for functional miRNA overexpression.	94
Table 3.5	Product codes of TaqMan® primer/probe assays used for miRNA expression analysis.	95
Table 3.6	Product IDs of TaqMan® primer/probe assays used for pluripotency and differentiation marker analysis.	95
Table 3.7	Primary antibodies for Western blot.	96
Table 3.8	List of miRNAs upregulated in NTera-2 cells upon 3-day exposure to 10^{-2} mM ATRA which potentially target crucial hESC markers.	117
Table 3.9	MiRNA expression in NTera-2 and 2102Ep cells exposed to 10^{-2} mM ATRA.	118
Table 3.10	Previously discovered functions of miR-10a, miR-99a and miR-184, miR-219 and miR-335.	119
Table 4.1	Volumes of transfection agent used for transfection optimisation and functional siRNA transfections in 2102Ep and NTera-2 cells.	165
Table 4.2	Different final concentrations of siRNA used for transfection optimisation in 2102Ep and NTera-2 cells.	165
Table 4.3	TaqMan® primer/probe assays used during siRNA transfection optimisation.	166
Table 4.4	Primary antibodies for Western Blot.	167
Table 4.5	Secondary antibodies for Western Blot.	167

Table 4.6	Product IDs and sequences of functional Silencer™ Select siRNA used in Chapter 4.	168
Table 4.7	TaqMan® primer/probe assays used during functional gene silencing.	169
Table 4.8	Primary and secondary antibodies for Western blot.	170
Table 4.9	Previous established functions of the five genes selected for silencing in 2102Ep and NTera-2 cells. In addition to functional information, the fold downregulation of these genes in NTera-2 cells exposed to 10^{-2} ATRA for three days is denoted.	189
Table 5.1	Transfection agents and volumes used for siSOX2 transfection in 2102Ep and NTera-2 cells.	225
Table 5.2	TaqMan® primer/probe assays used for gene expression analysis of SOX2 and GAPDH.	226
Table 5.3	Primary and secondary antibodies used for protein expression analysis of SOX2 and GAPDH.	227
Table 5.4	Visual quality control of the gene arrays.	229
Table 5.5	Computational quality control of the gene arrays.	229
Table 5.6	RNA purity and integrity in 2102Ep samples.	242
Table 5.7	RNA purity and integrity in NTera-2 samples.	242
Table 5.8	Differentially expressed genes in 2102Ep and NTera-2 cells.	252
Table 5.9	Top 10 upregulated and downregulated genes in 2102Ep cells.	258
Table 5.10	Top 10 downregulated and upregulated genes in NTera-2 cells.	258
Table 5.11	Genes commonly deregulated in both cell lines.	259
Table 5.12	Numbers of genes bound by OCT4/SOX2/NANOG as per Boyer et al. with the percentage fraction of each respective group.	262
Table 5.13	Numbers of genes that overlap with the data from Greber et al and this laboratory with the percentage fraction of each respective group.	262
Table 5.14	Numbers of genes that overlap with the data from Fang et al. with the percentage fraction of each respective group.	263
Table 5.15	MiRNA expression in 2102Ep and NTera-2 cells.	278
Table 5.16	Mean expression and data point distribution of miRNA expression in 2102Ep and NTera-2 cells.	278
Table 5.17	Deregulated miRNA in 2102Ep cells which overlap with miRNA expression results from Gallagher *et al.* and Fang et al. and the ChIP results from Marson *et al.* and Boyer *et al.*	282
Table 5.18	Deregulated miRNA in NTera-2 cells that overlap with miRNA expression results from Gallagher et al. and Fang et al. and the ChIP results from Marson et al and Boyer et al.	282
Table 5.19	Gene ontology of predicted targets of miRNA downregulated in 2102Ep cells.	287
Table 5.20	Gene ontology of predicted targets of miRNA upregulated in 2102Ep cells.	287

Table 5.21	The stated number of ≥2-fold deregulated genes which are predictably targeted by a stated number of ≥2-fold deregulated miRNA compared to the total number of ≥2-fold deregulated genes and miRNAs.	289
Table 5.22	Genes in 2102Ep cells targeted by a set of miRNA with a significantly deviating median.	294
Table 5.23	Genes in NTera-2 cells targeted by a set of miRNA with a significantly deviating median.	294
Table 5.24	Deregulated genes in 2102Ep and NTera-2 cells which have SOX2 binding sites in their promoters as defined by Boyer et al. and Fang et al.	302
Table 5.25	MiRNAs deregulated in 2102Ep and NTera-2 cells with previously documented functions in CSCs.	305
Table 5.26	Deregulated miRNAs in 2102Ep cells with enriched expression or documented functions in ESCs and ESC differentiation, including previously reported roles in cancer.	307
Table 5.27	Deregulated miRNAs in NTera-2 cells with enriched expression or documented functions in ESCs and ESC differentiation, including previously reported roles in cancer.	308
Table 5.28	Inversely regulated target genes and their associated miRNAs.	314
Table 5.29	Previously validated targets of miRNAs deregulated in 2102Ep and NTera-2 cells. Green miRNAs/genes were upregulated. Red miRNAs/genes were downregulated.	314
Table 5.30	Deregulated miRNA families and clusters in 2102Ep and NTera-2 cells. Green miRNAs were upregulated. Red miRNAs were downregulated.	316

List of Figures

Figure 1.1	Timeline of important discoveries linked to cancer stemness, CSCs and ESCs.	5
Figure 1.2	Four therapeutic methods for CSC-specific treatment.	14
Figure 1.3	Three approaches to cancer treatment.	15
Figure 2.1	The reverse transfection work-flow.	48
Figure 2.2	The forward transfection work-flow.	48
Figure 2.3	The miRNA TaqMan® qRT-PCR work-flow and chemical process.	63
Figure 2.4	TaqMan® Array card.	64
Figure 2.5	General gene array workflow.	73
Figure 2.6	SDS-PAGE gel transfer sandwich.	78
Figure 3.1	Cell culture plate layout for functional transfections of Pre-miR™ precursors.	94
Figure 3.2	Ct expression values of β-actin and XenoRNA™ versus cell numbers in 2102Ep and NTera-2 cells.	100
Figure 3.3	The comparison of PTK9 expression normalised to GAPDH expression in 2102Ep and NTera-2 cells.	102
Figure 3.4	Cell viability of 2102Ep and NTera-2 cells grown in Opti-MEM® I and hEC media.	104
Figure 3.5	Expression of PTK9 in 2102Ep cells reverse transfected with miR-1 three three different volumes of Lipofectamine™ RNAiMAX™, siPORT™ NeoFX™ and DharmaFECT 4.	106
Figure 3.6	Percentage expression of PTK9 in NTera-2 cells reverse transfected with miR-1 using three different volumes of Lipofectamine™ RNAiMAX™, siPORT™ NeoFX™ and DharmaFECT 4.	108
Figure 3.7	Cell viability of 2102Ep cells incubated in Opti-MEM® I for 6 hours versus cells exposed only to hEC medium.	110
Figure 3.8	Cell viability of NTera-2 cells incubated in Opti-MEM® I for 6 hours versus cells exposed only to hEC medium.	110
Figure 3.9	Percentage expression of PTK9 in 2102Ep cells forward transfected with miR-1 using three different volumes of Lipofectamine™ 2000.	113
Figure 3.10	Percentage expression of PTK9 in NTera-2 cells forward transfected with miR-1 using three different volumes of Lipofectamine™ RNAiMAX.	113
Figure 3.11	Cell viability of 2102Ep cells transfected with 50 nM pmNC#1 at different concentrations of Lipofectamine™ 2000.	115
Figure 3.12	Cell viability of NTera-2 cells transfected with 50 nM pmNC#1 at different concentrations of Lipofectamine™ RNAiMAX.	115
Figure 3.13	Morphology of 2102Ep cells transfected with Pre-miR-10a, -99a and -184 and incubated for three days.	122

Figure 3.14	Morphology of 2102Ep cells transfected with Pre-miR-219 and Pre-miR-335 and incubated for four days.	123
Figure 3.15	MiR-10a, miR-99a and miR-184 expression in functionally transfected 2102Ep cells.	125
Figure 3.16	MiR-219 and miR-335 expression in functionally transfected 2102Ep cells.	126
Figure 3.17	Pluripotency marker expression in 2102Ep cells transfected with Pre-miR-10a, -99a and 184.	128
Figure 3.18	Pluripotency marker expression in 2102Ep cells transfected with Pre-miR-219 and Pre-miR-335.	129
Figure 3.19	Differentiation marker expression in 2102Ep cells transfected with Pre-miR-10a, -99a and 184.	132
Figure 3.20	Morphology of NTera-2 cells transfected with Pre-miR™ precursors and incubated for three days.	134
Figure 3.21	Morphology of NTera-2 cells transfected with Pre-miR™ precursors and incubated for four days.	135
Figure 3.22	MiR-99a and miR-184 expression in functionally transfected NTera-2 cells.	137
Figure 3.23	MiR-219 and miR-335 expression in functionally transfected NTera-2 cells.	138
Figure 3.24	Pluripotency marker expression in NTera-2 transfected with Pre-miR-10a, -99a and 184.	140
Figure 3.25	Pluripotency marker expression in NTera-2 cells transfected with Pre-miR-219 and Pre-miR-335.	141
Figure 3.26	Differentiation marker expression in NTera-2 transfected with Pre-miR-10a, -99a and -184.	143
Figure 3.27	Morphology of NTera-2 cells transfected with Pre-miR™ precursors and incubated for 6 days.	145
Figure 3.28	MiR-335 and miR-219 expression in NTera-2 cells transfected with Pre-miR-335 and Pre-miR-219 (6 days after transfection).	146
Figure 3.29	Pluripotency marker expression in NTera-2 cells transfected with Pre-miR-335 (6 days post-transfection).	146
Figure 3.30	OCT4 and SOX2 protein expression in NTera-2 cells transfected with Pre-miR-335.	148
Figure 4.1	Cell culture plate layout of functional transfections of siRNAs.	168
Figure 4.2	SOX2 silencing in 2102Ep cells transfected with 10 and 15 nM siSOX2.	173
Figure 4.3	GAPDH silencing in NTera-2 cells transfected with 30 nM siGAPDH.	173
Figure 4.4	Western blot of SOX2 silencing in 2102Ep cells.	174
Figure 4.5	Cell viability of 2102Ep and NTera-2 cells transfected with ssNC#1 at three different concentrations of Lipofectamine™ 2000 and Lipofectamine™ RNAiMAX respectively.	176
Figure 4.6	Microscopic images of 2102Ep cells transfected with siSOX2 and ssNC#1.	178
Figure 4.7	Stemness marker expression in 2102Ep cells transfected with siSOX2.	179
Figure 4.8	Differentiation Marker Expression in siSOX2 transfected 2102Ep cells.	180

Figure 4.9	Staining of AP on 2102Ep cells transfected with siSOX2 and ssNC#1 and incubated for four and 7 days.	182
Figure 4.10	Staining of AP on NTera-2 cells exposed to 10^{-2} mM ATRA for four days.	182
Figure 4.11	Microscopic images of NTera-2 cells transfected with siSOX2 and ssNC#1.	184
Figure 4.12	Pluripotency marker expression in NTera-2 cells three days after transfection with 30 nM siSOX2.	185
Figure 4.13	SOX2 expression in NTera-2 cells transfected with 5, 10 and 15 nM siSOX2.	186
Figure 4.14	Pluripotency marker expression in NTera-2 cells at 6 days after transfection with 30 nM siSOX2.	186
Figure 4.15	Staining of AP on NTera-2 cells transfected with siSOX2 and ssNC#1 and incubated for four days.	187
Figure 4.16	Images of 2102Ep cells after transfection of siRNA targeting five genes followed by a 3-day incubation.	191
Figure 4.17	Gene expression in 2102Ep cells after transfection of siDPPA3, siSHISA2, siRBM47, siAK3L1 and siVSNL1.	194
Figure 4.18	Relative expression of DPPA3, SHISA2, RBM47, AK3L1 and VSNL1 in undifferentiated 2102Ep cells.	195
Figure 4.19	Western blots of silenced genes in 2102Ep cells.	196
Figure 4.20	Stemness marker expression in 2102Ep cells transfected with six siRNA.	197
Figure 4.21	Staining of AP on 2102Ep cells transfected with specific siRNA and untransfected cells.	199
Figure 4.22	Cell viability of 2102Ep cells four days after specific gene silencing.	201
Figure 4.23	Morphology of NTera-2 cells transfected with siRNA targeting the five genes and incubated for four days.	203
Figure 4.24	Morphology of NTera-2 cells transfected with siRNA targeting the five genes and incubated for 6 days.	204
Figure 4.25	Gene expression in NTera-2 cells after transfection of siDPPA3, siSHISA2, siRBM47, siAK3L1 and siVSNL1.	207
Figure 4.26	Relative expression of DPPA3, SHISA2, RBM47, AK3L1 and VSNL1 in undifferentiated NTera-2 cells.	208
Figure 4.27	Western blots of silenced genes in NTera-2 cells.	209
Figure 4.28	Stemness Marker expression in functionally transfected NTera-2 cells three and six days post-transfection.	211
Figure 4.29	Staining of AP on NTera-2 cells transfected with specific siRNA and untransfected cells.	213
Figure 4.30	Cell viability of 2102Ep cells four days after specific gene silencing.	214
Figure 5.1	SOX2 expression in 2102Ep cells three days after transfection with 15 nM siSOX2.	239
Figure 5.2	Western blot analysis of SOX2 expression in siSOX2 transfected 2102Ep cells.	239

Figure 5.3	Western blot analysis of SOX2 expression in siSOX2 transfected NTera-2 cells.	240
Figure 5.4	Electropherograms of single 2102Ep and NTera-2 RNA samples.	242
Figure 5.5	Example of electropherograms from RNA with different integrities.	243
Figure 5.6	Visual quality control of the GeneChip® Human Gene 1.0 ST array.	245
Figure 5.7	Different quality control metrics for gene array analysis of 2102Ep samples.	247
Figure 5.8	Different quality control metrics for gene array analysis of NTera-2 samples.	248
Figure 5.9	Relative Log Expression values in 2102Ep and NTera-2 samples.	249
Figure 5.10	Hierarchical clustering of 2102Ep and NTera-2 biological replicates.	250
Figure 5.11	Pearson's correlation of unmodified fluorescence.	251
Figure 5.12	Venn diagrams of deregulated genes overlapping between 2102Ep and NTera-2 cells.	252
Figure 5.13	Correlation of Ln-transformed gene array and qRT-PCR expression values from 2102Ep cells.	254
Figure 5.14	Correlation of Ln-transformed gene array and qRT-PCR expression values from NTera-2 cells.	254
Figure 5.15	The untransformed gene array and qRT-PCR expression values from 2102Ep cells.	255
Figure 5.16	The untransformed gene array and qRT-PCR expression values from 2102Ep cells.	255
Figure 5.17	Correlation of the 58 differentially expressed genes shared between the 2102Ep and NTera-2 datasets.	260
Figure 5.18	Deregulated genes in important signalling pathways in NTera-2 cells.	264
Figure 5.19	Deregulated genes in important signalling pathways in 2102Ep cells.	265
Figure 5.20	Most highly enriched functional annotation clusters in 2102Ep differential gene expression data.	268
Figure 5.21	Most highly enriched functional annotation clusters among upregulated genes in NTera-2 cells.	270
Figure 5.22	MiR-16 expression in NTera-2 cells using unamplified an preamplified cDNA.	272
Figure 5.23	qRT-PCR amplification plot of miR-16 expression in preamplified and unamplified 2102Ep cDNA.	272
Figure 5.24	MiR-16 expression in cDNA synthesised and preamplified from different quantities of 2102Ep RNA.	273
Figure 5.25	Box plots of Ct distribution gamut of expressed miRNA in 2102Ep cells.	275
Figure 5.26	Box plots of Ct distribution gamut of expressed miRNA in NTera-2 cells.	276
Figure 5.27	Ct signal correlation between biological replicates.	277
Figure 5.28	MiRNA differentially regulated in 2102Ep cells.	279
Figure 5.29	MiRNA differentially regulated in NTera-2 cells.	280

Figure 5.30	MiRNA differentially regulated in both 2102Ep and NTera-2 cells.	280
Figure 5.31	Frequency of genes as predicted targets in multiple miRNA in 2102Ep and NTera-2 cells.	286
Figure 5.32	Frequency of predicted miRNA-mRNA pairs per differentially regulated gene in 2102Ep cells.	291
Figure 5.33	Frequency of predicted miRNA-mRNA pairs per differentially regulated gene in NTera-2 cells.	292
Figure 5.34	Skewness of the distributions of expression values of predicted targets in 2102Ep cells.	292

Abbreviations

AML	acute promyelocytic leukaemia
AP	Alkaline Phosphatase
ATRA	All-trans Retionic Acid
BSA	Bovine Serum Albumin
CLL	chronic lymphocytic leukaemia
cDNA	Complementary DNA
ChIP	Chromatin Immunoprecipitation
ChIP-seq	ChIP sequencing
CNS	CNS
CSC	Cancer Stem Cell
CTC	Circulating Tumour Cell
DMEM	Dulbecco's Modified Eagle Medium
dsRNA	double stranded RNA
ECC	EC cell
ECL	enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
EMT	Epithelial-to-mesenchymal Transition
EpiSC	Epi Stem Cell
ESC	Embryonic Stem Cell
FBS	Foetal Bovine Serum
FDR	False Discovery Rate
FRV	Fast Red Violet
GO	Gene Ontology
hECC	human ECC
hESC	human ESC
HFT	High-frequency Target
hMSC	human MSC

HRP	Horse Radish Peroxidase
HSC	Haematopoietic Stem Cells
iPSCs	Induced Pluripotent Stem Cells
LRA	Luciferase Reporter Assay
LSC	Leukaemic Stem Cells
MCSC	migrating CSC
mECC	mouse ECC
mEF	mouse embryonic fibroblasts
mESC	mouse ESC
MET	Mesenchymal-to-epithelial Transition
MGB	Minor Groove Binder
miRNA	microRNA
MSC	Mesenchymal Stem Cell
MTT	Dimethyl Thiazolyl Diphenyl Tetrazolium Salt
NCBI	National Center for Biotechnology Information
ncRNA	non-coding RNA
NSC	Normal Stem Cell
NTC	Non-transfected Control
PBS	Phoshpate Buffered Saline
PBST	PBS-Tween
PCR	Polymerase Chain Reaction
Pen-Strep	Penicilin-Streptomycin
PM	Plasma Membrane
pmNC#1	Pre-miR™ Negative Control #1
PTU	Polycistronic Transcription Unit
PVDF	Polyvinylidene Diflouride
qRT-PCR	quantitative Real-Time-PCR
RAR	Retinoic Acid Receptor

RARE	RAR Element
RCT	Random Control Target
RIN	RNA Integrity Number
RISC	RNA-induced Silencing Complex
RLE	Relative Log_2 Expression
RMA	Robust Multi-Array Average
RQ	Relative Quantification
RT	Reverse Transcription
RT°	Room Temperature
RXRE	Retinoid X Response Elements
SDS	Sodium Dodecyl Sulphate
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
shRNA	short hairpin RNA
siRNA	small intefering RNA
snuRNA	small nuclear RNA
ssNC#1	Silencer™ Select Negative Control #1
ssRNA	single stranded RNA
TGCT	Testicular Germ Cell Tumour
TLDA	TaqMan® Low-density Arrays
UTR	Untranslated Region

Publications

Journal Articles

Gallagher, MF., Flavin, RJ., ElBaruni, SA., McInerny, JK., Smyth, PC., Salley, YM., **Vencken, SF.**, O'Toole, SA., Laois, A., Lee, MY., Denning, K., Li, J., Aherne, ST., Lao, KQ., Martin, CM., Sheils, OM., O'Leary, JJ. Regulation of microRNA biosynthesis and expression in 2102Ep embryonal carcinoma stem cells is mirrored in ovarian serous adenocarcinoma patients. *J Ovarian Res* **2**, 19 (2009). PMID: 20015364

Published Abstracts

Vencken, SF., Gallagher, MF., O'Leary, JJ. Upregulation of miR-184 in a testicular embryonal carcinoma cell line downregulates key stemness markers and upregulates differentiation markers. School of Medicine Post-graduate Research Day 2010.

Vencken, SF., Gallagher, MF., ElBaruni, SA., Martin, CM., Sheils, OM., O'Leary, JJ. MiR-335 is upregulated upon retinoic acid-induced differentiation of NTera-2 human embryonal carcinoma cell line and can itself induce differentiation. Rediscovering pluripotency: from teratocarcinomas to embryonic stem cells (Abcam and Peter Andrews)

Vencken, SF., Gallagher, MF., ElBaruni, SA., Martin, CM., Sheils, OM., O'Leary, JJ. Novel mechanisms of sox2 expression and regulation revealed in embryonal carcinoma. School of Medicine Post-graduate Research Day 2011.

Vencken, SF., Gallagher, MF., Martin, CM., Sheils, OM., O'Leary, JJ. Mechanisms of expression and regulation of SOX2 and its targets in two embryonal carcinoma cell lines. Irish Association of Cancer Research 2012.

Vencken, SF., Gallagher, MF., ElBaruni, SA., Martin, CM., Sheils, OM., O'Leary, JJ. MiR-335 is upregulated upon retinoic acid-induced differentiation of NTera-2 human embryonal carcinoma cell line and can itself induce differentiation. United States & Canadian Academy of Pathology Conference 2012.

CHAPTER ONE

INTRODUCTION

1 — Introduction

1.1 — The Stem Cell Theory of Cancer

Pierce (1974) pointed out that many cancers are caricatures of their tissues of origin.[1] Tumours seemed to contain all the normal characteristics of cell proliferation and cell differentiation, albeit in a highly aberrant and uncontrolled manner. During the evolution of the tumour, this level of disorganisation increases. However, the tumour tissue still retains a degree of heterogeneity and complexity similar to that of normal tissues. The development of tumour tissue has frequently been observed to be similar to the hierarchical manner of normal tissue development.

Research performed in the past two centuries has gradually culminated into the current stem cell theory of cancer.[2] While studying teratocarcinomas, 19th-century pathologists, such as Virchow, Paget and Récamier first noticed the histological similarities between cancer cells and cells of an embryonic origin.[3–5] Later, pathologists such as Cohnheim and Durante proposed that these cancers could possibly arise from dormant embryonic remnants in adult tissue, which were reactivated.[6,7] However, by the beginning of the 20th century, this "embryonic rest theory of cancer" was generally discredited, as can be exemplified by Bainbridge in 1914, who stated, "The congenital or embryonic theory of the origin of cancer has received no support whatever from the experimental and comparative investigations of recent times".[8,9] In the 1950s, the first successful human teratocarcinoma xenografts were performed in hamsters by Pierce *et al.*[10] This work also initiated a cascade of research, which would ultimately lead to the discovery and establishment of mouse embryonic stem cells (mESCs) and human embryonic stem cells (hESCs). By subcutaneously grafting cancer patients with cells from their own disseminated cancer, Southam & Brunschwig discovered in 1961, that tumour cells had a tumour initiating abilities in humans.[11] In 1963, Bruce *et al.* showed that only 1-4% of murine lymphoma cells transplanted in mice would form colonies on the recipients' spleens.[12] Similar studies, performed not long after, showed similar results to this.[13,14] Concurrently, with similar techniques, McCulloch & Till created the first clonal method to identify stem cells by injecting normal bone marrow cells into immunocompromised mice and observing the growth of small colonies on the spleen. They proposed that these were derived from the original cells and in 1963, their research containing the first outlined criteria of stem cells was published.[15,16] Later, McCulloch also contributed to research that would lead the establishment of "mouse myeloma tumour stem cells [sic]" in semi-solid media.[17] The simultaneous and often related discoveries of stem-like cells in normal and cancerous tissue during this period seeded the field of stem cell research. Starting in the late 1960s extensive research was performed on embryonal carcinoma cell (ECC) cultures, which were

a malignant subpopulation of cells isolated from teratocarcinomas. From work that began in 1953, Stevens was able to produce testicular teratocarcinomas in the 129 mice strain by grafting normal germinal cells from the genital ridge of inbred 129 mice that had a disposition to teratocarcinogenesis.[18] Not long after, Stevens managed to induce the growth of teratomas—benign and more differentiated counterparts of teratocarcinomas—by intratesticular transplantation of primordial germ cells into the 129 mouse strain.[19] Similarly, Kleinsmith & Pierce transplanted single ECCs, harvested from "neoplastic embryoid bodies", into mice and found that 11% of these single cells gave rise to teratocarcinomas consisting of various differentiated tissue types.[20,21] Thus, it was concluded that these ECCs were the multipotential stem cell component of teratocarcinomas.[21] These discoveries and others created the foundation upon which later the fields of cancer stem cell (CSC), adult stem cell and embryonic stem cell (ESC) research were established. During the late 1960s and early 1970s, the first murine ECC (mECC) lines were established in culture by Finch & Ephrussi and Kahan & Ephrussi.[22,23] Other researchers followed soon after, which led to the establishment of many experimental benign and malignant mECC lines of mixed differentiation potential, such as NF-1, P19, SIKR, F9 and many others.[24–33] Soon after, the first human ECC (hECC) lines were established *in vitro,* such as the TERA1, TERA2 and SuSa cells.[34,35] Later, the widely used NTera-2, 2102Ep, GCT27 and NCCIT cells were established, of which some were used for the investigation of ECC pluripotency, a property earlier ECC lines either lacked or was mistakenly overlooked during earlier research.[36–41] The pluripotent NTera-2 cell line, capable of extensive differentiation, and the nullipotent 2102Ep cell line, capable of only limited differentiation, were both used as model cell lines for this thesis. The research into ECCs later contributed to the original isolation of mESCs in 1981 by Evans & Kaufman—also achieved separately by Martin—and the original isolation of hESCs by Thomson et al in 1998.[42–44] From the 1950s to the 1990s, research into ECC lines was predominantly focussed on uncovering the mechanisms of stem cell pluripotency and the ESC mechanisms during embryonic development. However, the huge body of research into embryonal carcinoma and the established cell lines continue to provide much insight into stem cell component of carcinogenesis, the role of stem cells in the tumour microenvironment and in tumour proliferation, invasion and metastasis.

While much research was performed on the 'stemness' of teratocarcinomas and ECCs, other groups were focussing on the CSC hypothesis. During the 1960s, Potter, Pierce and others built on the previously mentioned embryonal-rest hypothesis by describing cancer as an arrest of stem cell maturation during tissue determination. During this time a consensus was beginning to form around a theory formulating that "undifferentiated stem cells which play [...] an important role in tissue renewal and which lack the overt manifestation of the differentiated tissue are the target in carcinogenesis" and that "oncogeny is blocked ontogeny" i.e. the combination of aberrant cell

proliferation *and* differentiation, rather than proliferation alone.[45,46] Based on these notions, a general hypothesis was proposed that malignant stem cells should contain all the genomic attributes of normal stem cells, but that the former had capacity to self-renew at the expense of the capacity to differentiate. It was further postulated that due to the similarity of normal and neoplastic stem cells, the latter were the causative agent of a tumour's ability to "usurp the logistics of the host" for the purpose of invasion and metastases in a similar manner in which normal embryo's usurp the logistics of the maternal organism, at her expense.[45] During this research the notions were formed that heterogeneous tumours contained populations of stem cells in small minorities and that permanent tumour regression could possibly be achieved if focused on eliminating the CSC component.[45,47] This was partially based on the discovery that tumour malignancy was removed when squamous cell carcinoma stem cells were differentiated into benign, well-differentiated squamous cells with the incapacity of forming tumours.[48] Embryonal carcinoma is a relatively rare disease, however, from the research in the field teratocarcinomas and their progenitors, ECCs, the hypothesis was formed that CSCs are perhaps the progenitors of carcinogenicity in general.[45,49] Teratocarcinomas and ECCs were considered an aberrant form of embryogenesis and perhaps an intrinsically abnormal form of parthenogenesis, which could be unique to certain types of germ cell tumours.[50] The discovery of cancer progenitor cells with stem cell-like features as the carcinogenic factors of other cancer types was required to solidify the CSC theory. As mentioned previously, the studies performed by Southam & Brunschwig and Bruce *et al.* showed a cell colony initiating ability of a only small fraction from a larger population of transplanted cancer cells. Further conjecture on the existence of CSCs was supported by the establishment of mouse myeloma and human myeloma cell cultures *in vitro*, which suggested 0.001% to 0.1% of all myeloma tumour cells were clonogenic.[17,51,52] Almost twenty years later, as technology developed to better characterise cells based on the expression of defined cell surface markers, the first conclusive evidence was produced for the existence of CSCs. In 1994, Lapidot *et al.* isolated murine leukaemic stem cells (LSCs) with a cell-surface phenotype of normal, uncommitted haematopoietic stem cells (HSCs) (CD34+/CD38-).[53] They successfully showed that these cells could reinitiate and maintain malignancy in immunocompromised mice, which were histologically similar to those of the donor, but that more differentiated cells with a surface phenotype of CD34+/CD38+ could not do this.[53] Not long after, this experiment was repeated and it was shown that these tumour-initiating cells had the capacity for differentiation, proliferation and self-renewal.[54] After these discoveries, many researchers followed on by isolation and characterising CSCs from multiple different cancers, such as cancers of the breast (first CSCs derived from a solid tumour), central nervous system (CNS), colon, head and neck, skin, ovaries and pancreas.[55–63] In some malignant tissues, such as prostate cancer, it

has proven more difficult to isolate CSCs. However, with novel improvements in cell and molecular biology and researcher's understanding of CSCs, huge progress is being made in search for stem cell components in these cancers.[64,65]

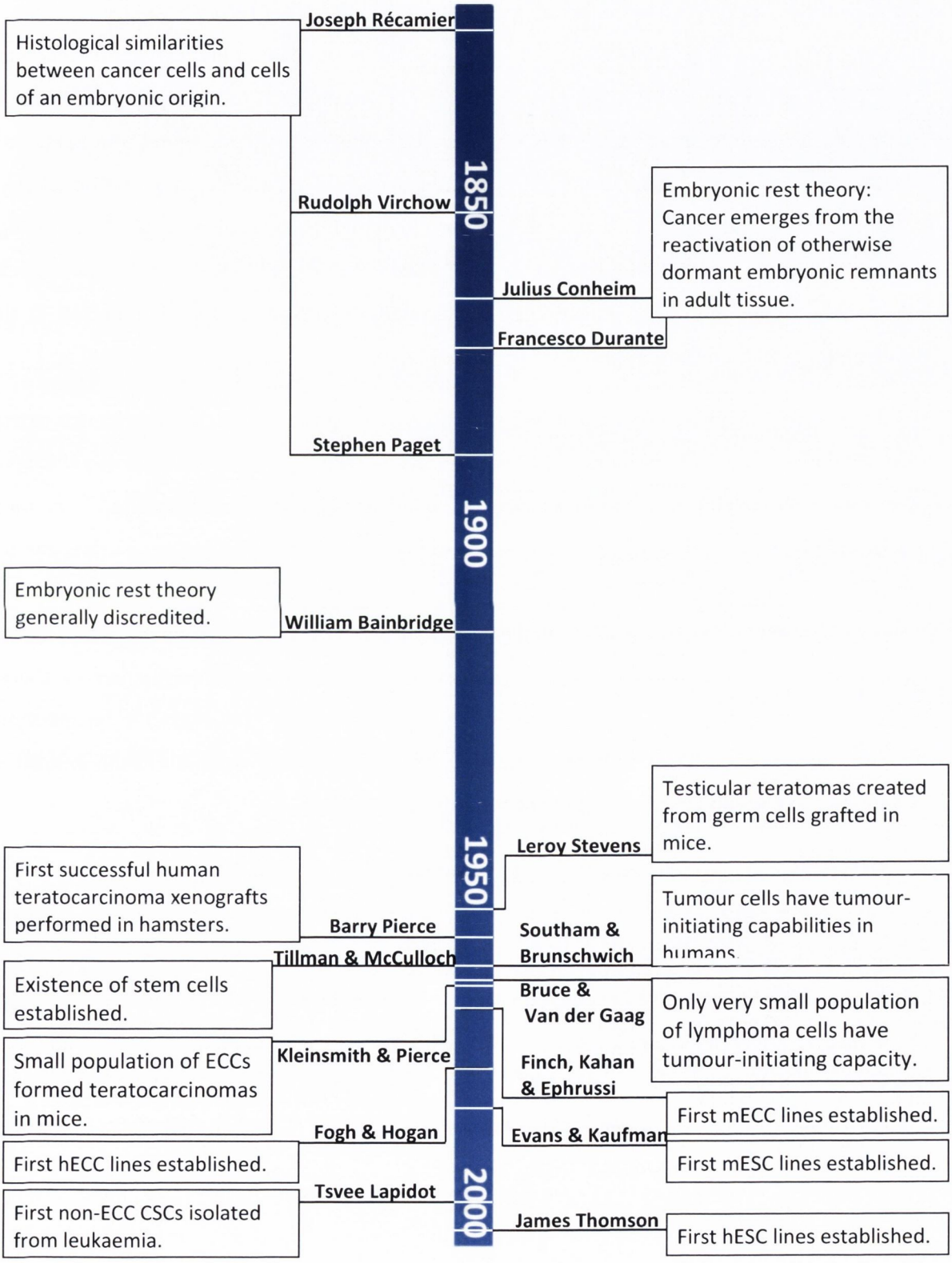


Figure 1.1—Timeline of important discoveries linked to cancer stemness, CSCs and ESCs. The individual lines represent a single year or short period along the timeline.

1.2 — The Role of Cancer Stem Cells in Carcinogenicity

Many researchers, such as Potter, have taken a holistic or "organistic" approach to cancer, by recognising that the disease is resultant from the interaction of cells and tissues within the proximal milieu of the tumour as well as being influenced by signals from more distal tissues. Others, such as Hunter, argued a reductionist view, which describes cancer as a consequence of intrinsic intracellular regulatory failure.[66,67] Ever since these approaches first evolved, researchers have produced much evidence to support both. For example, a reductionistic approach in molecular cell biology and genetics has revealed the numerous somatic mutations, epigenetic changes and oncogenic intracellular signalling pathways which contribute to an individual cell's malignant phenotype, while an organistic approach in pathology, physiology and tumour biology has yielded an understanding of the role of tissue homeostasis and the immunological, haematological and endocrinological roles in malignancy. This perspective is reflected in the developing theories on CSCs.[68]

From the beginning of CSC research, it was suggested that the biology of CSCs was similar to that of their normal stem cell counterparts (i.e. those which share the same developmental origin). CSCs are isolated from tissues based on the phenotype of their normal counterparts.[2,69] The most obvious similarities between these cell types are the ability to self-renew and proliferate, and the ability to differentiate into different cell types. Self-renewal and proliferation alone are not a valid argument for the existence of CSCs; malignancies in general rely on this ability. Many of the underlying mechanisms that govern the processes of self-renewal and differentiation are shared by normal stem cells and CSCs. Consequently, these mechanisms are considered to be important for carcinogenicity and have been the dominant focus of molecular research in the interest of understanding CSCs and in the interest of cancer therapy.

Phenotypically, normal stem cells and cancer stem cells are quite similar. A common method of isolating CSCs is with the aid of cell sorting techniques based of cell surface marker signatures. These signatures can comprise of one or multiple cell surface proteins, depending on the tissue of origin. These markers are mostly selected based on what is known about the development of the tissue type and they can be found to be expressed on the surface of normal stem cells from these tissues.[69,70] The two most commonly used surface markers used to identify CSCs are CD133 and CD44. Some of these cell surface markers are now potential targets of CSC-based cancer therapy.[71]

Both normal stem cells and CSCs share intracellular "stemness" pathways, many of which have been the focus of research into CSC-targeted therapy. Much of the research in the intracellular

pathways that govern normal and CSC behaviour has come from the study into HSCs and their malignant counterparts, LSCs.[2] Examples of important cancer stemness pathways and potential targets of CSCs-based cancer treatment are the Hedgehog/GLI, Notch, FGF, TGFβ/BMP, Wnt and PI3K/Akt signalling pathways.[71–73]

Besides intracellular signal transduction, common mechanisms of gene transcription can be found in normal stem cells and CSCs. Many transcriptional factors are, in fact, downstream targets of the aforementioned signalling pathways. Examples of these are the SMAD family, c-Myc and the β-catenin/LEF/TCF-related family.[71,73,74] In addition, cell cycle regulators such as p63 and the p53-related factors and the activation of telomere elongating factors (genes traditionally implicated in many cancer types) have been linked to the ability of CSCs to maintain self-renewal and ignore senescent signals.[75,76] An important family of genes in ESCs and ECCs are the pluripotency factors. Highly characterised key regulators of pluripotency, self-renewal and ESC maintenance are POU5F1 (OCT4), SOX2 and NANOG, but c-Myc, KLF4 and LIN28 among a core group are also considered important.[77–80] The members of this core pluripotency network directly co-regulate each other and autoregulate themselves to maintain a fine-tuned and balanced expression in their undifferentiated host cells.[77–80] This maintains the undifferentiated, self-renewing and, important for all hESCs and some CSCs, pluripotent state. The importance of these genes to pluripotency was highlighted during the production of the first induced pluripotent stem cells (iPSCs) from mouse and human somatic skin cells, which were almost indistinguishable from ESCs.[81–84] In these studies, a combination of these pluripotency factors were upregulated in somatic cells, producing iPSCs, which could differentiate in multiple different tissues. Interestingly, on some occasions, these cells produced cancerous lesions when injected into mice and the potential tumourgenicity of these cells is an implication for their use in regenerative medicine.[85] Histologically, poorly differentiated tumours (i.e. those with a high pathological tumour grade) show an overexpression of genes which are typically enriched in ESCs.[86] The aforementioned pluripotency factors are included in this set and they have been frequently implicated in cancer and CSC biology. The research performed on the functions of pluripotency factors in cancer, CSC biology and iPSC tumourgenicity have contributed to the idea that carcinogenicity and the generation of CSCs might involve the progressive reactivation of an ESC-like program proportionate to the aggressiveness of the disease.[78,87]

1.3 — Cancer Stem Cells and Metastasis

A characteristic feature of cancer is the ability to metastasise. This is the ability of cancer cells to undergo a metaplastic transition followed by a spread and invasion to other tissues, either proximal to the primary site or to a distant organ of the host. Cancer metastasis is responsible for 90% of cancer-associated mortality.[88] Briefly, the metastatic process of carcinomas consists of two phases: the first phase consists of the physical translocation of cancer cells by the lymph and blood vasculature to local or anatomically distant tissues, while during the second phase these cancer cells colonise these tissues and generate secondary metastatic lesions. During these phases, cancer cells undergo various transformations, which enables them to acquire an invasive phenotype and an adaptability to the different environments. The first such transformation is necessary for cells to break away from the primary tumour and gain the ability to migrate. Subsequent transformations enable the cells to invade local or distant tissues. The reason why some cancer cells do this so readily is still unclear. However, two possible working models have been suggested, both of which are not mutually exclusive. One model, the somatic-mutation theory of metastasis, describes cancer cells to acquire a sequence of metastatic and invasive traits by a Darwinian process, involving accumulation of somatic mutations, which increases the cancer cells' fitness with each clonal succession.[89] A more recent model suggests metastatic and invasive behaviour to be the effects of a local advantage for the cancer cell's survival. In other words, an epigenetically regulated "metastatic signature" can be found in the primary tumour and is linked to a cancer cell's other malignant properties such as uncontrolled self-renewal and proliferation.[90] Interestingly, gene expression studies have suggested that the predisposition of a tumour to metastasise could even be ingrained at its inception, possibly emphasising the importance of extrinsic signals in metastatic behaviour (i.e. the interaction of the metastatic cancer cell with the tumour microenvironment).[91]

For both these models, research is increasingly revealing the roles of CSCs in every step of metastasis. The ability of cancer cells to metastasise and initiate tumour growth is therefore associated with carcinogenesis, to which the somatic-mutation theory or epigenetic theory can apply. The ability of CSCs to self-renew and differentiate in to cells of different types gives them the preferential traits considered to be essential for the morphological transformations required for metastatic behaviour. Cells in both primary and secondary tumours have been shown to express stem cell markers from the tissue of origin.[92–95] Experimentally, the metastatic ability of CSCs becomes apparent when they are transplanted into an organism and allowed to reinitiate tumour growth. This also occurs when tumour-initiating cells are injected intravenously, showing the ability to survive in the circulatory system.[92,96] However, as most of these experiments are

performed in organisms with a compromised immune system, it has been argued that the microenvironment created by the active immune system of the original host, might, in reality show different *in vivo* results.[97] Other studies showed that cancer cells expressing particular cell surface markers associated with stem cells initiate tumours that themselves are more metastatic than those that do not. The self-adaptation of CSCs to survive an ectopic environment and gain the ability to invade dissimilar tissues, either those of the CSC donor or a different host, is a strong indicator that CSCs probably have the ability to do this during the second phase of metastasis. The following paragraphs will briefly describe research into the roles of CSC metastasis as published until the writing of this thesis.

Based on their research with colon cancer, Brabletz *et al.* (2005) proposed the existence of a subset of CSCs, which they called migrating cancer stem cells (MCSCs).[98] While most CSCs in the tumour are stationary, these MCSCs, located at the tumour-host interface are migratory and can metastasise. MCSCs have been characterised by their immunophenotype and by an alteration in global gene expression.[98,99] Although both stationary CSCs and MCSCs share typical stem cell markers, gene expression in stationary CSCs generally has an epithelial enrichment, while in MCSCs genes typically involved in the maesenchymal phenotype are enriched. Important for metastasis, the CSC phenotypes are readily interchangeable, enabling a transition and reversal of phenotypes necessary for dissemination and tissue invasion. Since this hypothesis was first proposed, various supporting papers have been published which describe the isolation of MCSC-like cells based on their immunophenotype, despite an apparent lack of specific migrating markers.[100–106]

Epithelial-mesenchymal transition (EMT) is widely recognised as an important step for efficient metastasis. This process is characterised by the trans-differentiation of cancer cells from an epithelioid to a meschymal or fibroblastoid morphology. EMT includes an array of processes, from cell-cell junction disassembly, actin cytoskeleton reorganisation, increasing cell motility and invasion of the basement membrane, which are governed by a regulatory change in specific cellular pathways, such as Wnt/β-catenin, Notch and TGFβ.[99] In 2007 and 2008, it was proposed to classify EMT as a process tailored to three distinct biological contexts.[107] Type 1 EMT occurs during implantation, embryogenesis and organ development. Type 2 EMT is associated with tissue regeneration and wound healing and Type 3 EMT is associated with cancer progression and metastasis. In the interest of this study, only Type 3 EMT will be discussed in this section, as it is relevant to this thesis.

The activation and inactivation of a number of important signalling pathways execute the EMT process. As discussed previously, the activation of Wnt/β-catenin, Notch and TGFβ is important

for EMT and can result in the transcription and repression of a large variety of genes. A much investigated step of EMT is the repression of the transmembrane cell-cell adhesion protein, E-cadherin, by Snail and Slug.[98] The consequential loss of cell-cell adhesion and the acquisition of a mobile state enables the morphogenically altered cancer cells to invade local tissues (local invasion) or to transform to circulating tumour cells (CTCs), which lympatically or hematogeneously travel to distant sites of invasion in the organism.[108] The existence of circulating malignant cells has been known for many years, but the clinical significance of detecting individual disseminated cells was questioned due to their inefficiency to seed secondary tumours (metastatic inefficiency), in contrast to circulating tumour residues (minimal residual disease), which are frequently detected and readily metastasise.[109] However, recently small subpopulations of CTCs have been discovered with potent metastatic abilities.[108] These cells can be of much smaller size than tumour residues, enabling deep tissue invasion by means of the microvascular system.[88] Up till now, the discovery of these cells was exceedingly difficult due to their extreme rarity compared to normal cells in the circulatory system.[108] Modern technology and improved isolation of CTCs orientated on their immunophenotype has facilitated the development and commercialisation of CTC-based clinical procedures, which show a substantial diagnostic potential.[108]

As mentioned previously, EMT is not unique to cancer and has initially been discovered to be a crucial process during embryonic development, during which epithelial cells are highly plastic and frequently alternate between epithelioid and mesenchymal states via EMT and mesenchymal-epithelial transition (MET).[110,111] The transition and reversal of these phenotypes are important for the metastatic process to complete, as a transformation back to an epithelial phenotype is required for the initiation of secondary tumours. Current hypotheses hold the broad notion that while EMT leads to the expression of stem cell markers, this cannot alone explain the phenotypic plasticity needed for optimal malignant behaviour, and that therefore EMT is more intrinsically linked to stem cell-like properties.[93] The convergence of MCSC-enabled metastasis and the process of cancer cell EMT is now a important topic of research, yielding a more complete picture of the, still quite elusive, metastatic cascade.

1.4 — Cancer Stem Cell-targeted Cancer Therapy and Its Implications

The CSC hypothesis states that a fraction of the total tumour cell population of a cancer consists of stem cell-like cells which fuel the propagation of the disease. This hypothesis also predicts that after the treatment of cancer with conventional therapies such as chemoradiotherapy, immunotherapy or surgery, a small population of CSCs can remain in the host, which can reinitiate tumour growth.[112] Current generation cancer therapies primarily focus on the elimination of tumours and cancer cells in an indiscriminate, non-targeted fashion. Surgery, immuno-, chemo- or radiotherapy are applied with the aim of completely removing malignant tissue from the body. They are often successful in removing a bulk, within a degree of probability, sufficient to avoid a relapse of the disease. Frequently however, after a period of time, the cancerous tissue can re-appear, sometimes more aggressively and often with an increased resistance to the therapeutic methods used.[113] The concept of cancer recurrence and resistance fits the CSC hypothesis well as tumour heterogeneity endows different cells with different abilities of resisting treatment and re-initiating tumour growth. The resistance of CSCs to common chemotherapeutic treatments is being supported by an increasing body of evidence and, besides their ability to re-initiate tumour growth, contributes further to their tumorigenicity.[114]

The natural ability to sequester somatic mutations by their high mitotic rate might give CSCs an evolutionary advantage to gain resistance to traditional chemotherapy or radiotherapy. However, recent research has shown that also the plasticity of CSCs and the nature of their stem cell phenotype endow them with a boosted ability to do this.[115] In addition, normal stem cells and CSCs are able to avoid therapies that target fast-dividing cells, by temporarily residing in a quiescent state.[116] Chemotherapeutic drugs are often designed to selectively target fast dividing cells. CSCs fall within this category. However, they employ a number of methods to resist chemotherapy. From relatively early on it was known that CSCs express high levels of ATP-binding cassette drug transporters, enabling CSCs to rapidly pump drugs out.[117] Besides this, CSCs exhibit an active DNA-repair system and resistance to apoptosis.[116] Furthermore, signalling pathways involved in stemness, such as Wnt/β-catenin and Notch have been implicated with chemo- and radioresistance by e.g. providing tolerance to DNA damage.[118] Recent research has shown that CSCs thrive in an hypoxic environment by the transcriptional activity of the hypoxia inducible factors, which can promote the CSC phenotype.[119] A tumour environment contains hypoxic regions in which CSCs thrive. Chemo- or radiotherapy can promote these hypoxic regions by creating a necrotic environment, in which, as a result, the CSC phenotype can be strongly promoted.[119,120] The hypoxic environment is one example of the microenvironmental niches which can contribute to CSC maintenance and resistance. [119] These aforementioned mechanisms inhibit

the efficacy of treatment that target these mechanisms and increases the ability of CSCs to survive chemo- and radiotherapy.

One of the hallmarks of cancer is its ability to evade destruction by the host's immune system. CSCs have been recently shown to play a critical role in this. The adaptation of the immune system to the tumour microenvironment facilitates the survival of the cancer.[97] Evidence has even shown that tumour cells can employ certain parts of the immune system for their own benefit, such as the recruitment of immune inflammatory cells to promote angiogenesis, proliferation and invasiveness.[114] There seems therefore to be two main provisions by which CSCs can function within the environment of the immune system. The first, and probably most important, is the ability of CSCs to avoid destruction by the immune system. The second is the CSC's ability to actively recruit the immune system for its own malicious purposes. Some mechanisms have been found by which CSCs can "hide" from the immune system. It has been known that cancer cells use "molecular mimicry" to hide their phenotype from the immune system by expressing a particular phenotype that mimics that of non-cancerous cells. In the past four years researchers have detected the presence of immunomodulatory cell surface markers on CSCs, such as CD200, which could possibly contribute to immune evasion.[121,122] Recently, LSCs have been found to overexpress the cell surface maker, CD47, to mimic the phenotype of normal HSCs and hide them from macrophages, indicating that CSCs can use the mechanism of molecular mimicry.[123] This research suggests that the innate stem cell phenotype of CSCs can possibly be responsible for immune system evasion. An increasing body of evidence is also suggesting that cells undergoing EMT display an increased ability for immune evasion and an increased resistance to immunotherapy based on the expression of certain proteins that facilitate this ability.[124] Other methods of immune evasion and suppression have been associated with CSCs. Cancer cells can produce a wide variety of immunosuppressive molecules and recent research has found this phenomenon in CSCs.[125] CSCs' ability to recruit the immune system has been less well characterised, but much progress is being made. It has been shown that interaction of breast cancer cells with the immune system can stimulate tumour EMT and appearance of cancer cells with stem cell-like phenotypes and it was found that CD8 T-cells could induce these cells to undergo EMT.[126] CSC immune evasion and the ability to recruit the immune system for carcinogenic purposes has severe implications for immunotherapy.[124] For forty years, blood-platelets have been known to facilitate cancer metastasis.[127] This phenomenon has been recently been revisited in the context of CSCs and their vascular adaptations, CTCs. These cells can interact with various haematological constituents, such as coagulation factors and blood platelets.[128,129] CTCs can recruit and activate these platelets and factors to avoid their elimination by the immune system while in the bloodstream.[130,131] The expression of CD142 (tissue factor) enables CTCs to bind platelets, aiding immunoevasion and

even diagnostic evasion by using platelets to 'cloak' their cell surface marker antigens.[88] This laboratory has shown a significant correlation between the extent of platelet recruitment and the metastatic potential of CTCs.[132] In addition to this function, various factors secreted by activated platelets and haemostatic factors can aid in the maintenance of a metastatic and pro-survival CTC microenvironment, while shielding CTCs from the otherwise hostile microenvironment of the bloodstream.[133] Haemostatic factors and platelets have been shown to be involved in nearly every component of the metastatic program, including tumour extravasation and tissue invasion, and are therefore highly researched therapeutic targets.[133]

The abilities of CSCs to resist commonly used forms of therapy and even modern molecular therapies originate from their plasticity, signalling redundancy and their dexterity to reinitiate the growth of resistant tumours. These are believed to be the primary reasons why treatment has been relatively unsuccessful for many different cancer types. Ideal cancer treatment regimen would eradicate both the bulk of differentiated cells as well as target the CSC component of the tumour, to avoid recurrence and resistance. Current research is focussed on CSC elimination and CSC differentiation or a combination of both as CSC targeted treatment. The first focuses on the elimination of CSCs themselves, while the second focuses on removing aspects of the CSC phenotype which contribute to cancer. For a large part, research is focussed on methods of identification of CSCs in different cancer types so a specific treatment can be tailored to target these cells. CSCs with prognostic relevance can be found by the expression of their molecular signature in much the same way as they characterised experimentally.[69] Overlap exists in treatment strategies for CSC elimination and CSC differentiation as the mechanisms of stemness maintenance in CSCs overlap with their survival. A recent review proposes four strategies for eradicating liver CSCs, which probably can be adapted to CSC treatment in general.[134] These are: the blocking of CSC pathways, ablation of prospective markers, disruption of the microenvironment and disruption of self-protection (see Figure 1.2 for a summary). Overlap exists among these strategies; self-protection (resistance) mechanisms of CSCs are linked to the pathways responsible for their phenotype (discussed earlier), the microenvironment is partially responsible for the preservation of self-protection mechanisms and regulation of CSC pathways and *vice versa*. Ablation therapy towards prognostic cell surface markers by e.g. medicinal antibodies targeting these markers or by targeting specific signalling events important to CSCs with a particular immunophenotype is a possible strategy and potential therapies targeting cells with CSC markers such as CD133 and EpCam are in development.[116,134,135] For this therapy to be specific, the CSC phenotype has to be unique. Unfortunately, current methods cannot yet guarantee CSC phenotype specificity *in vivo*, although promising advances are being made.[113] In

the interest of this study, the following paragraph will only focus on blocking CSC pathways and the removal of the CSC phenotype.

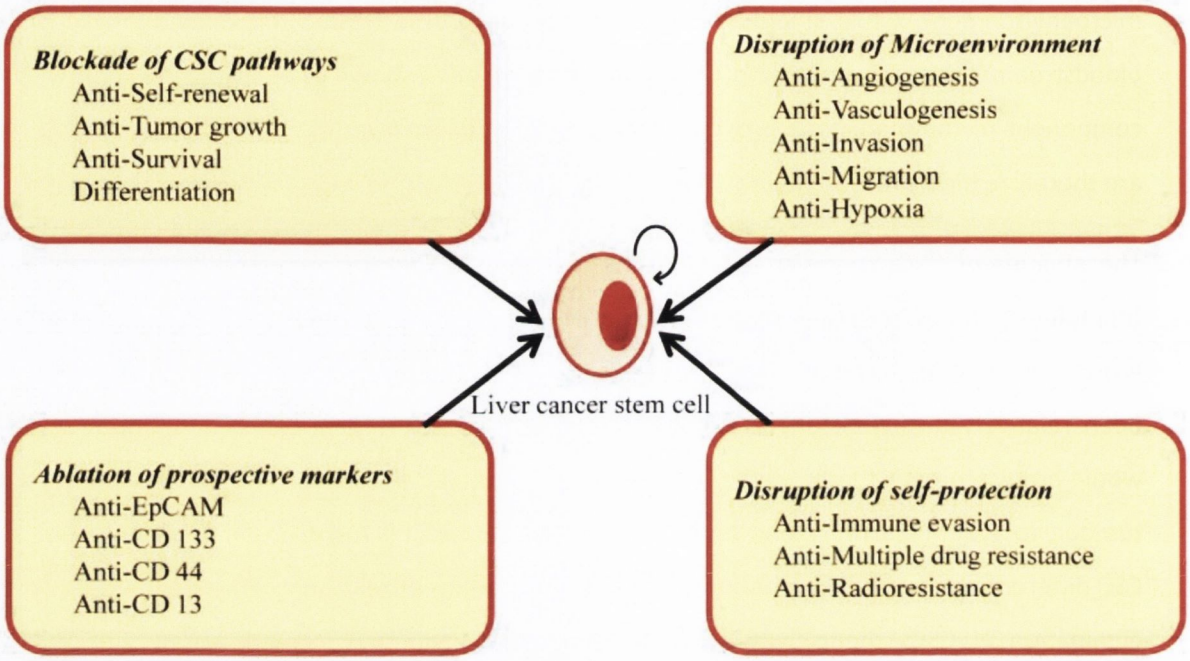


Figure 1.2—Four therapeutic methods for CSC-specific treatment. *(Ref. Oishi and Wang.[134])*

Targeting key CSC pathways is an approach that has garnered much interest. The focus lies on the removal of the CSC phenotype, resulting in 'benign' cells with no capacity to propagate tumour growth and spread. The targeting of these pathways can result in the removal of a particular property of CSCs, such as their ability to self-renew, or can result in a more broad morphological event such as differentiation. The Wnt/β-catenin is an important pathway for the maintenance of normal stem cell and CSC self-renewal and proliferation. This pathway has been proposed as a therapeutic target and recently, studies have been performed in which the inhibition of upstream regulators of the pathway, such as the Frizzled family of receptors or inhibiting or degrading more downstream elements, such as β-catenin lead to the reduction of carcinogenicity.[136–143] Also included as a treatment option is the induced upregulation of endogenous inhibitors of the Wnt/β-catenin pathway, such as DKK1.[144,145] The Hedgehog pathway has also been suggested as a potential therapeutic target.[134] The inhibition of the Hedgehog pathway by the induced modulation of its regulatory factors has been shown to sensitise cancer cells to drug-induced apoptosis and sessility.[146,147] The implication of targeting specific pathways is that not all pathways are regulated in similar ways in all cancers.[148] Treatment options have also been explored with the focus on CSC differentiation, a more dramatic morphological alteration which could be a broader method of rendering them benign. This is possibly a potent method by which malignancy could be

eradicated. This is exemplified by recent research which showed that the potassium ionophore, salinomycin, could differentiate breast cancer CSCs, leading to cell-cycle arrest rather than the induction of cytotoxicity.[105] Salinomycin proved to be >100-fold more effective in removing breast cancer CSCs in mice than the common chemotherapeutic drug, paclitaxel. The TGFβ-related pathways, which are important for stem cell fate, are often underexpressed in CSCs and the induction of this pathway has been postulated to be an important contribution to differentiation treatment of cancer.[147] However, other studies have suggested the inhibition of particular TGFβ pathway elements can contribute to a reduction of metastasis, indicating that this pathway has diverse roles in cancer.[149] Retinoic acid is a potent morphogen which has long been known to be able to induce differentiation in ESCs and ECCs.[150] Retinoic acid has also been shown to differentiate cancer cells expression CSC phenotypes and is currently the most examined and clinically tested compound for differentiation treatment in cancer. Differentiation therapy by retinoic acid and its molecular analogues has already led to varied success being applied to treat acute promyelocytic leukaemia (AML), squamous cell carcinoma, adenocarcinoma of the prostate, neuroblastoma and hepatocellular carcinoma.[9] Because of its ability to stimulate differentiation, retinoic acid has also been proposed as a chemopreventative drug against cancer. Some research shows that retinoic acid could change what is otherwise a microenvironment that stimulates pre-malignant cells to an environment that encourages their differentiation.[9] More recent research has demonstrated how retinoic acid affects CSCs. Retinoic acid is investigated as a treatment for glioblastoma as it has the ability to reduce the expression of neural stem cell markers in glioblastoma stem cells, such as CD133, MSI1, nestin and SOX2 and differentiate these cells into senescent astrocytes.[151,152] Similar modulatory performance of retinoic acid has also been show in breast cancer stem cells.[153] Another reported advantageous effect of CSC differentiation is the potential subsequent exposure of previously hidden CSCs to the immune system, further ensuing their eradication.[154,155]

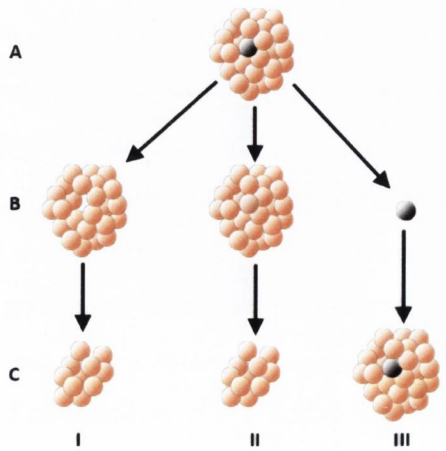


Figure 1.3—Three approaches to cancer treatment. The tumour contains a bulk of differentiated cells (orange) with at the root a small population of CSCs (black) (A-II). Conventional cancer treatment eliminates the bulk of tumour, while failing to completely eradicate the CSC population (B-III). The CSC reinitiates a tumour and a more resistant and malignant tumour recurs. Alternatively, targeted therapy can focus on the eradication or induced differentiation of the tumour's CSC component (B-I and B-II respectively). This can result in a permanent reduction or complete elimination of the tumour mass (C-I and C-II).

1.5 — The Origins of Cancer Stem Cells and the Heterogeneous versus Stochastic Model of Cancer

Despite the large number of publications referring to CSCs, a full consensus has not yet been reached on whether experimental evidence gained, such as that from CSC xenografting, reflects the real mechanisms of cancer as a disease.[156] This possibly questions the reliability of some CSC models and also questions the congruity of CSCs as a defined subclass itself. Although theory is progressing towards consensus about the heterogeneous nature of tumours containing stem-like cells, the theories on the origin and ultimate definition of these cells as CSCs *in vivo* are still controversial.[156] Two main hypotheses exist on the origin of CSCs, both of which could be correct. One hypothesis holds that cancerous cells with a stem-like phenotype originate from normal stem cells that have, by microenvironmental or intrinsic changes, such as oncogenic mutations, acquired a malignant phenotype. Another hypothesis holds that microenvironmental or intrinsic changes promote the stem-like phenotype to the much more abundant non-stem-cells or committed progenitor cells, which subsequently climb the hierarchy of cell potency and become malignant to produce heterogeneous tumours.[157,158] These hypotheses are frequently called the hierarchical and stochastic models of cancer respectively. A recent review positions the hierarchical cancer stem cell model against the stochastic clonal evolution model and tries to clearly define the differences between these two models.[157] Table 1.1 describes the differences as defined here, however the reviewers do concede that both models are not mutually exclusive in cancers that follow a stem cell model. Based on recent discoveries, many of which are described in this introduction, a reconciliation of these two models seems to be occurring as an increasing number of cancer types are found to have a stem cell component while the potential neoplasticity of seemingly more differentiated tumour cells is further elucidated and the cancer stem cell exclusivity is increasingly put into question.[159–162] For both models, convincing experimental evidence exists; CSCs have been shown to share large number of cell surface markers with their non-malignant counterparts and NSCs can adopt a malignant phenotype *in vivo*.[158] On the other hand, committed progenitors have also been shown to acquire a plastic, malignant phenotype by specific oncogenic transformations.[158] Strong evidence comes from the field of melanoma research. The tumour-initiating component of malignant melanoma was observed to be comparatively very large from which the hypothesis was held that the component of stem-like cancer cells in melanoma is maintained at a dynamic equilibrium, which can vary depending on environmental conditions.[157,163,164] A third, less commonly held hypothesis is that CSCs are the product of a rare fusion event between stem cells and other cell types.[115] It is possible that CSCs originate from this event or that this occurs after first-line treatment, producing recurring

CSCs.[115,158] It was postulated that cell fusion provides the resultant CSCs with the ability to acquire a phenotype necessary for oncogenic proliferation and for resistance to treatment. Indirect experimental evidence supports the cell fusion hypothesis as a facilitating event for initial CSC formation as for CSC recurrence.[115,158,165]

Table 1.1—Cancer stem cell model versus stochastic cancer model. *(adapted from Shackleton et al.[157])*

	Cancer Stem Cell Model	**(Stochastic) Clonal Evolution Model**
Frequency of cancer cells with tumorigenic potential	Rare to moderate	High
Phenotype of cancer cells	Heterogeneous	Heterogeneous or homogeneous
Tumor organization	Hierarchical	Not necessarily hierarchical
Intrinsic differences between tumorigenic and non-tumorigenic cells	Stable, epigenetic	Unstable, epigenetic or genetic
Rational approach to therapy	Possible to target only tumorigenic cells	Target most or all cells
Compelling clinical evidence	Germ lineage cancers	High-grade B cell lymphoblastic leukemia

1.6 — MicroRNA

MicroRNA (miRNA) are short RNA molecules with an average length of 21-22 nucleotides and are endogenous to eukaryotic organisms. MiRNAs are members of a superfamily of non-coding RNAs (ncRNAs). These RNAs in this superfamily, although dissimilar in length, share the attribute which is their incapability to undergo translation into proteins. Besides miRNA, some other members of this family include ribosomal RNA, small nucleolar RNA, small interfering RNA (siRNA), long ncRNA, transfer RNA and Piwi-interacting RNA. These RNA molecules have diverse cell regulatory or metabolic functions. The predominant function of miRNA in the cell is the post-transcriptional repression of mRNA translation (this is described in Section 1.7). The first miRNA was discovered in 1993 by Ambros, who revealed that a short anti-sense RNA expressed in *C. elegans* , dubbed lin-4, could post-transcriptionally downregulate its target gene, lin-14, without showing a reduction of this gene's mRNA expression.[166] In 2000, a second miRNA, let-7, was found in *C. elegans* by Reinhart *et al.* followed by the [167,168] Since then, thousands of miRNAs have been discovered in many different organisms and, currently, the Sanger MicroRNA Database holds approximately 1500 human miRNA sequences, although the number of unique mature miRNA sequences is lower.[169] MiRNAs are expressed in all cells of the body, however the miRNA profile is often cell type-specific.[170]

1.7 — Biosynthesis and molecular processing of MicroRNA

About 40% of all miRNA loci in the human genome are located in the intronic regions of protein coding genes and are often transcribed as a single polycistronic transcription units (PTUs). A much smaller number is located in the exonic regions of protein-coding genes; however this depends on alternate splicing patterns. Approximately 50% of all miRNAs are thought to be clustered in PTUs, either coding or non-coding, and are often governed by a single promoter region. In some cases, however, these miRNAs also have their own promoters.[171]

The biogenesis of mature miRNAs is fairly straightforward and starts with the transcription of an miRNA precursor RNA by RNA polymerase II, although in some cases RNA polymerase III adopts this role for miRNAs associated with Alu repeats.[172] This primary miRNA (pri-miRNA) can be several kilobases in length and forms double-stranded RNA (dsRNA) stem-loop structures containing a 5' cap and a poly-A tail which can contain the precursors of multiple mature miRNAs.[171] The pri-miRNA is cleaved in the single-stranded RNA (ssRNA) flanking regions of the stem-loop structures. This is performed in the cell nucleus by the Microprocessor complex, which contains Drosha, an RNase III-type protein, and Pasha, which assists Drosha in cleaving ssRNA segments approximately 11 base pairs away from the ssRNA-dsRNA junction.[171] This liberates a stem-loop constructs of approximately 60-70 base pairs called pre-miR.[173] The Microprocessor complex can cleave intronic pre-miRNAs without negatively affecting normal mRNA splicing, indicating an intricate regulatory link between this complex and the spliceosome. In some cases the Microprocessor is bypassed and intronic pre-miRNAs are produced by the spliceosome. These are sometimes called mitrons.[174] With the aid of a Ran protein, the pre-miRNA is subsequently exported into the cytoplasm by the nuclear transport receptor, exportin 5, which recognises stem-loop and 3' overhang of the pre-miRNA.[171] In the cytoplasm the terminal loop of the pre-miRNAs are cleave by the cytoplasmic RNase III, Dicer, which also contains a PAZ domain that binds to the 3' protruding end of the miRNA. The length of the dsRNA miRNA:miRNA* product depends on the overall stem-loop length and loop size, as well as any imperfect pairing in the stem. In some cases, as with mRNA isomers, a reference sequence can produce multiple mature variants of a single miRNA by post-transcriptional editing. These miRNAs are sometimes referred to as isomirs.[175] The mature miRNA:miRNA* duplex undergoes further processing as described in Section 1.8.

1.8 — Post-Transcriptional Regulation of Protein Synthesis by MiRNA

1.8.1 — The RNA-induced Silencing Complex

The primary cellular apparatus responsible for gene repression by miRNA is the RNA-induced silencing complex (RISC). This generic complex is a loose heterogeneous quaternary structure of a diverse composition of interchangeable proteins, which can load specific short regulatory dsRNAs for the specific silencing of genes at the transcript or genome level.[176] Although the mechanisms of the RISC and types of small RNA it can load are quite diverse, this Section will focus on RISC-mediated silencing of mRNA by miRNA. The RISC contains a number of different proteins with specific functions. The catalytic centre of the RISC are the evolutionary ubiquitous Argonaute proteins. This family of proteins can be arranged in three subgroups: the AGO group, the Piwi group and the WAGO group.[173] For miRNA-mediated silencing in mammals, the AGO group are important and specifically AGO1 and AGO2 (from a total of AGO1-4). Both AGO1 and AGO2 load miRNA, however, generally, only AGO2 loads both siRNA and miRNA and is capable of target mRNA cleavage.[173,177,178] Therefore, AGO1 is generally responsible for mRNA translation inhibition and AGO2 for mRNA degradation or processing, although exceptions to this exist as shown in *D. melanogaster* cells.[179] The sequence of RISC assembly depends on the configuration of proteins used. The miRNA loading process into the RISC is still quite obscure and differs depending on the RISC configuration (e.g. AGO1 or AGO2). Perfectly matched hybrids (miRNA or siRNA) are preferentially loaded in AGO2 and duplexes with central bulges (mostly miRNA) are loaded into AGO1.[180,181] In both cases the miRNA is loaded as a cleaved pre-miR dsRNA consisting of a guide strand and a passenger strand, after which the passenger strand is discarded and degraded in a RISC type-dependent manner.[173,181] In some cases, the passenger strand (denominated with a suffixal asterisk) escapes degradation and remains biologically active.[182–184] The loading of AGO2-RISC is dependent on a protein complex containing Dicer-2, however the AGO1-RISC loading mechanism is probably Dicer-independent, but is still rather obscure.[181,185] Both AGO1 and AGO2 contain a PAZ domain, which interacts with the 3' end of the guide strand and a PIWI domain, which interacts with the 5' end of the guide strand, orientating the miRNA for mRNA targeting.[176] RISC contains other important components, such as helicases, RNA binding proteins, deadenylases and regulatory proteins. An example of another critical component of RISC is the GW182 (*D. melanogaster*) family of proteins, which recruit the RISC components and mRNA to P-bodies for miRNA-induced silencing of genes.[186] Recent research has implicated GW182 with additional tasks within the RISC.[187] The loaded RISC is called the mature miRISC. Although not the only mechanism of mRNA repression, the most studied method of miRISC-mediate mRNA repression is by binding the 3'UTR of the mature mRNA.[188] This process is described in more detail in Section 1.8.2. The

intracellular location and general method of mRNA repression has been debated. Evidence has shown that mRNA repression can occur before, during or after the initiation stage of mRNA translation.[188] Furthermore, mRNA degradation or storage can occur in defined cytoplasmic locations, called P-bodies, or can remain at the polysomal level, or even both simultaneously. The RISC and mature miRNAs have even been found in the nucleus, suggesting a possibility of pre-mRNA repression or transcriptional repression.[189,190]

1.8.2 — Targeting mRNAs

The main function of miRNA is the post-transcriptional repression of mRNA. MiRNAs are believed to target up to 60% of expressed mammalian protein coding genes.[191,192] Supported by target prediction software and microarray analysis studies, miRNAs are widely accepted to each target tens to hundreds of genes.[193] In addition, it has been experimentally confirmed that one mRNA can be targeted by multiple miRNAs simultaneously.[194] The silencing of mRNA translation by miRNA is governed by a complex mechanism that is not fully understood. Over the past ten years a few "rules" that somewhat elucidate this mechanism have been discovered. The perfect Watson-Crick base pairing of seed sequence of the miRNA, a sequence between positions 2-7 from 5' end, is considered to be very important for target specificity and stable alignment, although exceptions do exist and, as of yet, no solid mechanistic theory explains this importance.[195,196] Also, efficient formation of the miRNA:mRNA duplex depends on the thermodynamic stability of the resulting hybrid, rather than perfect complementarity.[195] More recently it was confirmed that the thermodynamic binding process is slightly more complex, as miRNA target specificity also depends on the thermodynamic stability of any intramolecular hybridisation of the mRNA near the binding site.[197] This means that total free-energy change of miRNA:mRNA binding depends on both the free-energy gain of miRNA:mRNA hybridisation and the free-energy cost of opening the mRNA secondary structure.[198] Beyond the seed sequence imperfect complementarity can exist, but the degree of this can affect mRNA repression. Reasonable complementarity in the 3' region of the miRNA is required (especially in the 13-16 nucleotide region) and some mismatches or bulges are permitted. It has been shown that an adenosine across from position 1 of the miRNA or an adenosine or uracil across from position 9 improves site efficiency even if base paring does not occur.[188] Initially, it was thought that only sequences within the 3'UTR of the mRNA was targeted by the 5' end of the miRNA. However, recent research showed that also the coding region and the 5' untranslated region (5'UTR) could be targeted for mRNA silencing.[199–201] Similar to the function of siRNA, the silencing of mRNA by miRNA can occur by cleavage often resulting in the irreversible degradation of the mRNA. Some of the cleavage products have been shown to be biologically stable, suggesting that they might have

further functions in the cell.[196] The initially discovered and common function miRNA is the temporal repression of the mRNA, with the absence of degradation. In this case, only protein translation is repressed and the mRNA could possibly be reactivated for translation at a later time point. This has been theorised to be a transcriptional timing mechanism that could be reversed over time or during a cellular event.[188,202] It has been shown that a bulge or mismatch within the central region of the miRNA:mRNA hybrid generally precludes the cleavage of the mRNA, but this is not always the case.[196] The miRNA and their 3'UTR miRNA binding regions of mRNA are often evolutionary conserved; this is not a prerequisite for *in vivo* mRNA silencing by miRNA, increased conservation has been shown to positively impact apparent repression effictiveness.[203] Nonetheless, some studies showed that among experimentally validated binding regions, almost half were not conserved.[193] A generally accepted rule is that the degree of mRNA repression by a specific miRNA increases proportionally to the number of its target sites found on the mRNA 3'UTR.[196] So far, the rules that have been discovered to govern miRNA:mRNA hybridisation produce hundreds of potential targets for a single miRNA. When exceptions to these rules are included, this number will grow exponentially. The number of targets might not reflect miRNA function *in vivo*. Results from miRNA target validation experiments in cell cultures often show an expected phenotypic change associated with the downregulation of a single or handful of miRNA targets when the expression of a specific miRNA is artificially changed. From this, one can infer that perhaps an experimental justification exists for the idea that a single miRNA might predominantly target fewer genes than predicted by prediction algorithms that abide by the current rules, as only a few targets are needed for the desired or expected phenotypic change. The dynamics of miRNA:mRNA hybridisation are very complex and much is still unclear. Currently, research in the role of secondary structures and site accessibility of miRNA and mRNA and their complexes is clarifying some of additional methods by which miRNAs 'pick' their targets. Another important field of investigation is the transcriptional and post-transcriptional regulation of miRNAs themselves. This is discussed briefly in Section 1.9.

1.9 — Regulation of MicroRNA

Like genes, miRNA are regulated by a variety of epigenetic and post-transcriptional processes. As described in Section 1.7, miRNA loci exist in intergenic and intragenic (predominantly intronic) regions of the genome. Of these miRNA most intergenic miRNA are autonomously regulated by their own promoters, while intragenic miRNA are generally regulated by the promoter regions of their host genes, as is evident from their correlated expression patterns.[204] Interestingly, gene ontology has shown that co-expressed intronic miRNA and host genes cooperatively regulate similar cell processes.[204] The promoter regions of autonomously expressed (mostly intergenic) miRNAs are highly similar to those of protein-coding genes, with the presence of CpG islands, TATA box sequences, initiation elements and histone modifications.[179] These miRNA are therefore controlled in a similar manner. P53 is an important transcription factor and cell cycle regulator and its gene TP53 is mutated or deleted in over 50% of all cancers.[205] The reduction of p53 function is also well documented in CSCs and regulates the transcription of important tumour suppressor miRNAs, with often critical functions in CSCs, such as the miR-200 family (further details on miR-200 function in CSCs are described in Section 1.12).[206] Auto-regulatory loops also exist, in which miRNAs are transcribed or repressed by the transcription factors they simultaneously regulate. A good example of this is the direct repression of pluripotency factors, OCT4, SOX2 and KLF4 by miR-145 in hESCs and the coexisting repression of miR-145 by OCT4, which binds its promoter.[207] As a consequence, miRNA expression can be carefully regulated by epigenetic mechanisms, such as methylation patterns and histone modification. An example discussed in more detail in Section 1.11, describes the silencing of miR-27 by hypermethylation in bladder cancer, which promotes the disease.[208]

Regulation of miRNA expression can also be governed by the miRNA processing machinery, such as the Microprocessor complex. The regulation of the components of this complex, Drosha and DGCR8, are intimately linked with one-another. It has been shown that DGCR8 stabilises the Drosha protein through protein-protein interaction and that Drosha mRNA can be reciprocally cleaved by Microprocessor.[209] These interactions will ultimately have an effect on global miRNA expression. Global miRNA regulation can also occur at the pre-miRNA processing level. For example, the stability of TRBP, an important protein in Dicer-mediated pre-miRNA processing, is positively regulated by the MAPK/ERK pathway.[179] Thus miRNA maturation is regulated by this pathway. More specific regulation of miRNAs by proteins also exist. The best example of this is the regulation of the let-7 family by LIN28. This protein can regulate these miRNAs at multiple levels by binding to both pri-let-7 and pre-let-7, inhibiting their further processing to mature let-7, a feature important for ESC maintenance.[179] Interestingly, LIN28 mRNA is itself a target of the let-

7 family. Other similar examples exist and are often regulated by upstream pathways. In Section 1.7, isomirs were described to be multiple different miRNAs derived from a single reference sequence. This is an example of miRNA editing, a regulatory method which can have an impact on miRNA target specificity. Another example of this is conversion of adenosine bases to inosine by adenosine deaminases.[179] Pri-miRNAs and pre-miRNAs can be edited by this mechanism, which can ultimately affect downstream miRNA maturation.

At the functional level, miRNAs can be regulated as components of the miRISC. Among the components of the miRISC, AGO and GW182 proteins are shown to be heavily targeted for regulation.[179] The regulated levels of AGO protein in the cell have been shown to not only affect miRNA-mediated target repression quantitatively, but possibly also qualitatively as different miRNAs may differ in their intrinsic ability to be loaded into the miRISC.[179] In mammals, multiple splice variants of GW182 proteins are expressed and the regulation by (de)phosphorylation of these proteins have been suggested to be a regulatory mechanism for miRNA function.[179] MiRNA regulation independent from RISC exist, such as (temporary) recruitment by RNA-binding proteins and by RNA sequences, such as pseudogenes, that can function as endogenous miRNA sponges.[179,210] Even the direct inhibition of a miRNA by another miRNA has been shown.[211] The existence of these alternate functions and the possible existence of others, indicates the miRNA network to be more complex than initially thought.

1.10 — Additional Functions of miRNA

Although the role of post-transcriptional gene suppression remains dominant, a comprehensive collection other functions have been attributed to miRNAs. These functions can promulgate beyond the cytosol and mature miRNAs have been found to exert their influence in the nucleus of the cell, but also in the extracellular domain. Besides their role to suppress protein translation, miRNAs have been found to activate gene expression, both at the transcriptional level, by binding to gene promoter regions, as at the post-transcriptional level, by binding to the 5'UTRs of mRNAs.[212,213] Furthermore, transcriptional inhibition of genes by miRNA has also been reported.[190] MiRNAs can transverse the plasma membrane and extracellular domain and enter other cells, by exosomal secretion.[214] This may represent a novel intercellular signalling mechanism.

1.11 — MicroRNA and Cancer

Shortly after the characterisation of miRNA in 2000, the implications of these short oligonucleotides in cancer biology were suggested and these roles of miRNA in cancer have been a highly investigated subject since.[215] It is estimated that approximately half of known miRNAs in the genome are located within or close to genomic fragile sites and regions involved in cancer.[216] It is now well known that miRNAs play important parts in carcinogenicity, tumour proliferation, metastasis and many other biological processes that can either directly or indirectly affect cancer. However, still relatively unknown is how crucial the role of miRNAs are in the creation of the carcinogenic state, as many other cell regulatory processes beyond miRNA regulation are involved. Interestingly, global miRNA expression in cancer has been demonstrated to be lower than that in normal tissue.[217,218] Similar to protein coding genes, miRNAs can be deregulated in a manner that is often characteristic to the cancer type and sometimes can be associated with the prognosis and progression of the disease.[219] One of the first miRNAs to be discovered to have a defined role in cancer was miR-21, which was upregulated in breast cancer and regulates the expression of BCL2.[220] Since then, hundreds of miRNAs have been associated with regulation of known and potential oncogenes and tumour suppressor genes. MiRNA which have a tumour suppressive function are often called tumour suppressor miRNA and those with an oncogenic function are called oncomirs. A typical attribute of many cancers is the often specific downregulation of tumour suppressor miRNAs and upregulation of oncomirs. As with other genes, the loss of tumour suppressor miRNA function can be due to various mechanisms such as genomic deletion, mutation, epigenetic silencing, altered transcription factor function and alterations in the miRNA processing machinery.[219] A common example of this is the deletion of the 13q14.3 region in chronic lymphocytic leukaemia (CLL), which contains miR-15a and miR-16, resulting to an accelerated proliferation of CLL due to their loss and subsequent upregulation of their targets, BCL2 and c-MYB oncogenes.[221] Fewer miRNAs have been classified as oncomirs than tumour suppressor miRNAs, however those that have been defined are well researched. Since miR-21 gained its oncomir status in CLL, this miRNA has been associated with many other types of cancer, such as cancer of the gastrointestinal tract, lung, bladder, cervix, pancreas, breast, head and neck, liver, prostate and leukaemia, lymphoma and melanoma.[218,222–231] Among its validated targets are important tumour suppressors, for example PDCD4, TGFBR2, PTEN and TIMP3.[232–234]

As described in Section 1.9, miRNAs can be regulated in various ways, with the most prominent being the regulation of their transcription by various transcription factors. MiRNAs play a role at every level of cellular signalling pathways, including those that play a role in cancer. Through miRNAs, transcription factors can exert an additional layer of influence on cell signalling.

Transcription factors with important roles in cancer have been found to transcribe oncomirs or tumour suppressor miRNAs. A good example of this is the active repression of the let-7 and miR-29 families of tumour suppression miRNAs by the oncogene c-Myc.[235] Specific hyper- or hypomethylation of CpG islands in miRNA promoter regions have been shown to be a common method of regulating miRNA expression. In cancer, methylation patterns are dramatically altered and as a consequence, this can lead to the upregulation of oncomirs by promoter hypomethylation or, more often, the transcriptional silencing of tumour suppressor miRNAs by hypermethylation.[219] For example, miR-127 was one of the first miRNAs discovered to be silenced by hypermethylation in cancer, specifically bladder cancer. This miRNA was further validated to target BCL6, a confirmed oncogene.[208] Examples of other tumour suppressor miRNAs that have been found to be silenced by promoter hypermethylation in various cancers are miR-124a and, more recently, miR-152.[236,237] Consequently, their induced upregulation resulted in reduced malignancy and the suppression of their target oncogenes. Some oncomirs have been found to be upregulated due to promoter hypomethylation. An example of this is let-7a-3, whose promoter was hypomethylated in lung adenocarcinomas and subsequently was highly overexpressed in this cancer.[238] This unusually oncogenic miRNA (most let-7 members are tumour suppressor miRNAs), has regulatory functions in the Ras pathway based on microarray data, although target validation still has to be performed. Very recently, in NTera-2 ECCs the hypermethylation of gene introns have been shown to contribute to the downregulation of the tumour suppressor, miR-199a.[239]

The association of the deregulation of specific miRNAs with cancer is generally performed empirically, by first isolating the sought cancer phenotype and subsequently evaluating the expression of miRNAs in the experimental (cancer) phenotype with a control phenotype (non-cancerous counterparts). A difference in expression of some or many aberrantly expressed miRNAs are subsequently inferred to contribute to the cancer phenotype and often this presumption is validated by returning the cancerous phenotype to that of a (more) normal phenotype by induced deregulation of these miRNAs. In some cases an alternative, more hypothetical approach to elucidating miRNA function in cancer is taken in the form of a functional screening assay of miRNA. This generally involves the systematically induced deregulation of specific miRNAs in cells followed by the subsequent association of expected cancerous phenotypes with specific exogenously influenced miRNAs.[240] The discovery of the tumour suppressive properties of miR-21 is an example of an miRNA found by this approach.[240] Functional miRNA screening assays are often performed with a large number of miRNAs and is sometimes combined in modified form with the empirical approach when the results of the latter method yield a large quantity of deregulated miRNA. The functional screening assay is not unique in

cancer research and is also used in the detection of important genes that can contribute to certain examined phenotypes.[241]

The specificity of miRNA expression profiles to certain types of cancers at certain stages of the disease has major diagnostic and prognostic values. MiRNA profiling has even been suggested to be more accurate and consistent than gene profiling as a method to determine malignancy.[217] Genome-wide miRNA expression profiling techniques, such as miRNA microarrays and quantitative real-time-PCR (qRT-PCR) arrays, have the ability to differentiate cancer tissues from their normal counterparts. MiRNA expression remains quite stable in cancer tissue storage conditions, such as fresh frozen tissues and fixed-formalin paraffin embedded tissues, expediting their use as diagnostic markers.[240] MiRNAs have also been found to be stably expressed in body fluids and recent research has shown the expression profiles in these fluids can be altered by cancer.[242] Many cancers have now been associated with specific body fluid miRNA expression profiles and commercial development of sensitive diagnostic tools is in its early stages.[242,243]

Because of the important implications of miRNA with cancer, the therapeutic value of miRNA-based treatment has also been investigated. A key advantages of miRNA-based therapy is the ability of miRNAs to simultaneously and efficiently regulate the expression of oncogenic or tumour suppressive genes and signalling pathways exogeneously.[244] With this, a precise command of these genes is possible. As miRNAs are involved in malignancy and treatment resistance, miRNA-directed treatment could potentially remediate the causal aberrancies.[245] The often-subtle and temporal nature of miRNA-mediated gene repression can potentially reduce undesirable side-effects, which could otherwise appear with more aggressive approaches. Also, miRNA-based therapy is not limited to restoring aberrant miRNA expression to non-cancerous levels, but can also be targeted at miRNAs which are normally expressed, but could still contribute to resolving the disease.[246] Treatment which revolves around employing miRNA's therapeutic potential can be categorised into direct and indirect miRNA regulation. In the first category, most current research in miRNA-based treatment is broadly directed at the development and delivery of molecules that mimic miRNAs or antagonise their expression.[246]

Table 1.2 shows a selection of different, potentially therapeutic methods that can achieve these functions. Indirect methods of miRNA-based treatment involve targeting the mechanisms that govern miRNA expression. As discussed in Section 1.9, many upstream signalling pathways can regulate miRNA expression. The upstream receptors or more downstream factors of these pathways could be targeted to indirectly control the expression of specific miRNAs. As with other molecular-based therapies, this probably relies quite heavily on sufficient knowledge about the pathway's signalling and transcriptional destinations, so to impede unwanted side-effects and

facilitate specific signalling. A recent example showed that the upregulation of anti-metastatic miR-335 in breast cancer could be induced by the modulation of the nuclear aryl hydrocarbon receptor by its ligands, TCDD and MCDF, resulting in the downregulation of the metastatic gene, SOX4.[247] As described in the earlier, a loss of miR-15a and miR-16 in CLL resulted in the overexpression of anti-apoptotic factor, BCL2. In a different variant of leukaemia, AML, the expression of these miRNA can be restored in AML cells and patients by administering all-trans retinoic acid (ATRA), resulting in the repression of BCL2.[248] The development of direct miRNA-based therapeutics is complicated. As with other RNA interference (RNAi)-based therapies, *in vivo* delivery of miRNA mimics or antagonists can be problematic due to the size of the molecules. However, miRNA-targeted delivery is considered easier than siRNA delivery due to the smaller size of the oligonucleotides (or analogues).[244] Various viral and non-viral strategies have been devised to overcome the problems of oligonucleotide degradation, immunogenicity, tissue and cellular uptake and non-specificity. The use of viral vectors for *in vivo* DNA delivery is a well-researched method for gene therapy and recently it has been used for *in vivo* miRNA delivery in murine liver cancer.[249] Viral vectors could additionally be used to deliver miRNA inhibiting sequences such as miRNA sponges (antisense oligonucleotides with multiple miRNA binding sites). Non-viral vectors are also being tested, such as cationic lipids, nanoparticles and polymers.[244] Indirect targeting might be easier as small-molecule inhibitors, which can potentially inhibit the miRNA transcription and processing machinery, and drugs affecting upstream signalling pathways show less difficulty in penetrating tissues and, if necessary, entering the cytosol and nucleus.[244] The potentially lower specificity of this type of treatment can be favourable, if target diversity is desired, or potentially disadvantageous if adverse off-target effects are not expected.

The exponential increase in miRNA research has firmly established the roles of miRNA in cancer, although much is yet to be uncovered. With the continued development of experimental tools to analyse miRNA regulatory networks in cancer, the increased application of such tools to cancer diagnosis and the potentials of miRNA-directed therapy, the attention to miRNA in molecular-based medicine is becoming increasingly accepted.

Table 1.2—Structural characteristics miRNA-base therapeutics and delivery methods. *(Adapted from Bader et al.[246])*

	Gene therapy	miRNA mimics	siRNAs	miRNA antagonists	Small-molecule inhibitor
Therapeutic					
Type of therapeutic	Protein	Short ssRNA	Short ssRNA	Short ssRNA	Small organic compound
Application	Replacement	Replacement	Inhibition	Inhibition	Inhibition
Status of drug target	Loss of function	Loss of function	Gain of function	Gain of function	Gain of function
Recognition by body as	Native	Native	Foreign	Foreign	Foreign
Delivery					
Structure of API	Large DNA vector	Short ds RNA	Short ds RNA	Short ss RNA	Small organic compound
Molecular weighta	~22 000 kDa	~15 kDa	~15 kDa	~7 kDa	~0.5 kDa
Delivery route	Local	Local/systemic	Local/systemic	Local/systemic	Systemic
Endogenous processing into final therapeutic	Transcription translation	RISC	RISC	None	None

1.12 — MiRNA regulation in Cancer Stem Cells and Embryonal Carcinoma

MiRNAs are very important in virtually every aspect of self-renewal, proliferation and differentiation processes of stem cells. This role has have been found in ESCs, germ line stem cells and somatic tissue stem cells and has been validated e.g. by the deletion of Microprocessor proteins, causing a reduction in ESC proliferation and a stalling of these cells in the G1 phase of the cell-cycle and also severe defects in ESC differentiation in mice.[250–252] Between differentiated embryoid bodies and ESCs, more than 100 miRNAs are differentially expressed.[175] Only very recently are the roles of miRNAs in CSCs beginning to emerge. Due to their comparable phenotype, miRNA regulation is probably quite similar in normal stem cells and CSCs. Gene regulation and cell signalling in normal stem cells is to some extent mirrored in CSCs. As discussed earlier in this introduction, many of the common regulatory mechanisms that govern stem cell maintenance, proliferation and differentiation have been linked to malignant behaviour. Global miRNA expression in poorly differentiated tumours is lower than in more differentiated ones.[217] This might suggest that miRNA expression is suppressed in CSCs compared to more or fully differentiated cells as more undifferentiated tumours are likely to contain a higher proportion of cancer cells with a stem-like phenotype. This is also a feature of ESCs, which, in their undifferentiated state, generally express relatively few miRNAs.[253] The low expression of let-7 and miR-200 family members in many different cancer types have previously been linked to ESCs and now also associated to CSCs.[254–258] For let-7, this was further validated by the induction of this miRNA in breast CSCs, which resulted in a suppression of tumour initiation in a mouse model.[256]

Interestingly, LIN28, a pluripotency factor, is strongly linked with negatively regulating let-7 expression, further substantiating the overlap between CSCs, stemness and miRNA.[259] Interestingly, when miR-200c was introduced to both breast CSCs and ECCs, this reduced the expression of BMI1, a gene critical for stem cell self-renewal, and resulted in the suppression of the breast CSCs and the differentiation of ECCs.[254] Another example is the c-Myc-activated miR-17-92 cluster of miRNAs, which has been shown to be an oncogenic polycistron in various cancers.[260] The expression pattern of this miRNA cluster is shared between CSCs and ESCs and has important functions in both.[261,262] A significant discovery is the ability of miR-17-92 miRNAs to promote self-renewal and differentiation evasion of LSCs by post-transcriptionally suppressing the potent cyclin-dependent kinase inhibitor, p21.[263] In P19 and NTera-2 ECCs, the miR-302 cluster was linked to the undifferentiated state of these cell lines.[264] It was shown that the members of this cluster downregulated upon the differentiation of these cells and that they are important for the maintenance of their stemness.[265] Shortly after these discoveries, the association of miR-302 with the undifferentiated state of hESCs was established and the miR-302 miRNAs were shown to be regulated by the OCT4/SOX2 dimer.[266] Very recent research has functionally linked other aberrantly expressed miRNAs the CSC phenotype, such as the CSC inhibiting miRNAs, miR-199-5p, -150, -451 and miR-34a, but also CSC stimulating miR-21, -9/9*, -17 and miR-181.[234,258,267–272] In fact, the epigenetic regulation of miR-34a has even been experimentally targeted in pancreatic CSCs for potential therapeutic purposes by restoring its expression with demethylating agents.[269] Further potential links between CSCs and miRNAs are currently being investigated.[273] As of yet, no miRNAs that differentiate between CSCs and normal stem cells have been found, however the therapeutic potentials for CSC-targeted, miRNA-based treatments are promising.

Some research has been conducted to establish links between miRNAs and embryonal carcinoma. This laboratory has performed differential whole-miRNAome analysis studies in pluri- and nullipotent ECC lines exposed to retinoic acid, revealed the contribution of miRNA to the resistance of 210Ep cells to differentiation.[274] Similar experiments in NTera-2 and P19 ECCs showed that the ESC-related miR-302 cluster downregulated during differentiation and the neural miR-124 upregulated.[264] Another study showed that the co-culture of ECCs with normal prostrate stromal fibromuscular cells contributed to the appearance of a cancerous miRNA and mRNA profile in these latter cells.[275] The regulation of the miR-302 cluster by crucial pluripotency factor, OCT4, has established another link of this gene to mECC self-renewal.[276] The low expression of OCT4 and miR-106b has been associated with high p21 and high cisplatin resistance in human nullipotent 2102Ep hECCs.[277] The downregulation of miR-383 and miR-199a promotes proliferation of human pluripotent NTera-2 ECCs and a restoration of miR-199a expression led to the suppression of cell growth, cancer migration, invasion and metastasis.[239,278] Chapter 3 of this

thesis aims to contribute to further elucidating the function of miRNA regulation in embryonal carcinoma.

1.13 — The Cancer Stem Cell Model of this Study

The cancer stem cell models used for this study are two hECC lines, 2102Ep and NTera-2. The 2102Ep cell line was derived from a testicular germ cell tumour (TGCT).[41] This cell line is resistant to ATRA-induced differentiation and will show truncated trophectodermal differentiation when cultured at low density.[279] The parental cell-line of the NTera-2 cell line was obtained by Peter Andrews from the University of Sheffield from a nude mouse xenograft of the Tera-2 cells.[37] This cell line undergoes differentiation along neuroectodermal lineages when grown at low density, which can be assisted by ATRA.[36] As detailed in previous Sections, hECCs are the CSCs of teratocarcinomas and it can be said that these cells are the archetype CSC, because of the extensive characterisation of their carcinogenicity and stem cell phenotype.[157] These cell lines have been previously used as cancer stem cell models. As described earlier, they also have much resemblance with hESCs and they express common undifferentiated hESC surface markers such as alkaline phosphatase, SSEA3, SSEA4, TRA-1-60 and TRA-1-81.[280,281] They, furthermore, express crucial pluripotency markers, OCT4, SOX2 and NANOG.[36,280] In many types of germ line cancer cells, such as ECC, pluripotent capacity is reactivated, which distinguishes these cells from their normal germ cell counterparts.[282] These properties make hECCs good models to study the overlap of CSCs, hESCs and pluripotency. Additionally, 2102Ep and NTera-2 cells do not require feeders to grow, which other are generally required by hESCs.[36] In addition, NTera-2 cells differentiate extensively to distinct cell fates in response to ATRA, HMBA and the bone morphogenetic proteins.[36] These properties make 2102Ep and NTera-2 cells relatively easy to culture and facilitate differentiation experiments. 2102Ep and NTera-2 growth patterns are quite similar to hESCs; they tend to grow in tightly packed colonies containing many small cells with large nuclei.[36] Both 2102Ep and NTera-2 cells, but in particular the former cell line, have a highly aneuploid chromosomal complement.[281] Interestingly, relatively recent research has shown some hESCs to contain aneuploid chromosomes and phenotypically resemble hECC lines, among which 2102Ep cells.[283] This observation further supports the close similarities of hESCs and hECCs.

Table 1.3—The two CSC models used in this study.

NTera-2	2102Ep
Pluripotent cells with a capacity to differentiate into different tissue types	Nullipotent with a limited ability to differentiate
Sensitive to ATRA-induced differentiation	Do not differentiate upon exposure to ATRA
Mostly forms well-differentiated tumours in mice	Forms highly undifferentiated tumours in mice

1.14 — Hypotheses

Based on pre-existing research, the premises of this study is that CSCs form the hierarchical apex underlying the genesis and malignancy of cancer and that the malignancy of cancer is inversely proportional to the differentiated state of its CSCs. CSC differentiation therapy has successfully been used before and the presupposition of this study is that CSC differentiation is a key method for eliminating malignancy. The hypotheses of this study hold that, as abnormal regulation of genes and miRNAs can contribute to the aberrant undifferentiated state of CSCs, targeting these mechanisms can cause CSCs to differentiate by which potentially therapeutic targets could be revealed.

The aims of this study were:

- To induce differentiation in ECCs by induced deregulation of specific miRNAs.
- To induce differentiation in ECCs by induced deregulation of specific genes.
- To elucidate the genes and miRNAs regulated by key pluripotency factors.

CHAPTER TWO

MATERIALS AND METHODS

2 — Materials and Methods

This Section lays out the general methods used throughout the thesis. Methods of similar application are grouped into Sections. For most methods, the background methodology is described before the exact protocol is explained.

2.1 — Cell Culture

2.1.3 — Cell Lines

Two testicular embryonal carcinoma cell lines were used for this study:

1. **2102Ep:** a human adherent, malignant nullipotent ECC line. The cell passage numbers of the 2102Ep cells used in this study were kept within the 35-50 range. The 2102Ep cells used in this study were a kind gift from Peter Andrews.
2. **NTera-2 cl.D1** or **NT2/D1**: a human, adherent, malignant, pluripotent ECC line. The passage numbers of NTera-2 cells used in this study were kept within the 40-50 range. The NTera-2 cells used in this study were a kind gift from Peter Andrews.

2.1.4 — Initial Cell Seeding

For both cell lines, frozen 1 ml samples from a passage number between 40 and 50, with a total cell number of approximately 10^6 to 1.5×10^6 cells were thawed at 37°C for < 1 minute and transferred directly to a 25 cm^2 cell culture flask *(83.1810.002, Sarstedt AG & Co.)* containing 6 ml of hEC medium at 37°C. hEC medium was prepared from Dulbecco's Modified Eagle Medium (DMEM; *BE12-604F, Lonza Group Ltd.*) with 4.5 g/L glucose and L-glutamine *(BE12-604F, Lonza Group Ltd.)* supplemented with 10% foetal bovine serum (FBS) *(DE14-801FG, Lonza Group Ltd.)* and 2% Penicillin-Streptomycin (Pen-Strep) *(DE17-602E, Lonza Group Ltd)*. After an overnight culture the medium was refreshed and the cells were left to grow to full confluence in humidified conditions at 37°C and 5% CO_2, replenishing with pre-warmed hEC medium every 1-2 days.

2.1.5 — Cell Culture Maintenance

Both cell lines were maintained at high density (>67,000 cells/cm^2) in 75 cm^2 cell culture flasks *(83.1813.002, Sarstedt AG & Co.)* in humidified conditions at 37°C and 5% CO_2. When the cells were at full confluence, they were passaged at no lower than 1:3. Cells were passaged by the following method:

- NTera-2 cells were passaged by scraping with a cell scraper *(83.1830 or 83.1832, Sarstedt AG & Co)* in hEC medium. The cell suspension transferred to a 15 ml falcon tube. The remaining cells were washed off the flask surface with 1X phosphate buffered saline (PBS) *(BE17-516F, Lonza Ltd)*. After a 3-minute centrifugation at 1000 × g, the supernatant was removed and the cell pellet was resuspended in a volume of pre-warmed hEC medium. A desired fraction of cell suspension was transferred to a 75 cm^2 cell culture flask.
- 2102Ep cells were washed with 1X PBS and detached by incubation in pre-warmed 0.25% trypsin-EDTA *(25200072, Invitrogen, Life Technologies)* for 5-10 minutes. The cell suspension was transferred to a falcon tube and an equivalent volume of hEC medium was used to collect the remaining cells and transfer them to the tube. After a 3-minute centrifugation at 1000 · g, the supernatant was removed and the cell pellet was resuspended in a volume of pre-warmed hEC medium. A desired fraction of cell suspension was transferred to a 75 cm^2 cell culture flask.

Cells cultured on a 6-well cell culture plate were split 1:2 at defined time-points. Cells were washed with 1 ml 1X PBS (37°C). NTera-2 cells were split by scraping and suspending on-plate in 1 ml hEC medium (37°C). 0.5 ml suspension was subsequently transferred to a new plate containing 2.5 ml hEC medium (37°C). 2102Ep cells were detached by 1 ml 0.25% trypsin-EDTA and transferred to a 1.5 ml microtube. 0.5 ml hEC medium was used to wash down remaining cells and this was also transferred to the same tube. The suspension was centrifuged at 1000 · g and the supernatant was removed. The cell pellet was resuspended in 1 ml hEC medium (37°C) and 0.5 ml was transferred to a new 6-well plate with 2.5 ml hEC medium per well.

2.1.6 — Long-Term Cell Storage

Cell freezing and storage was performed in similar manner for both cell lines. Cells were grown to full confluence in a 75 cm^2 flask. They were washed, harvested and pelleted as described in Section 2.1.5. The cells were subsequently resuspended in 6 ml Recovery™ Cell Culture Freezing medium *(12648-010, Invitrogen, Life Technologies)* and 1 ml aliquots were transferred to cryogenic storage vials *(368632, Nunc, Thermo Fisher Scientific)*. The cell suspensions were frozen overnight at -80°C under controlled conditions and subsequently stored in liquid nitrogen for long term storage.

2.1.7 — Cell Counting

For cell counting confluent NTera-2 and 2102Ep cells were harvested by scraping and trypsinisation respectively as described in Section 2.1.5. The cells were pelleted by centrifugation and diluted in 4 ml hEC medium. Full resuspension was effected by pipetting the cells with a sterile 3.5 ml Pasteur pipette. In two sterile 1.5 ml microtubes the cells were diluted further; 2 x 10 µl NTera-2 cell suspension was diluted in 10 µl hEC medium (37°C) or 2 x 10 µl 2102Ep cell suspension was diluted in 30 µl hEC medium (37°C). To the diluted cell suspensions an equivalent amount of sterile Trypan blue dye *(15250-061, Gibco, Invitrogen)* was added. This yielded two microtubes containing an NTera-2 cell dilution of 1:4 or two microtubes containing a 2102Ep cell dilution of 1:8. The cells were incubated for three minutes at room temperature and 10 µl from each microtube was added to each of two counting chambers of a haemocytomter *(Z35, 962-9, Sigma-Aldrich)*. Cells within four of the nine 1 mm^2 squares of the haemocytometer were counted in both chambers using a light microscope at 100X magnification. The total number of cells in counted in both chambers was averaged and subsequently multiplied by the dilution factor divided by the volume of the chamber counted, which is 4×10^{-4} cm^3. This yielded the concentration cells expressed as cells per ml in the original suspension.

2.1.8 — Cell Viability Assay

All cell viability and proliferation testing was performed with the MTT Cell Proliferation I kit *(11465007001, Roche Applied Science)*. The kit contains dimethyl thiazolyl diphenyl tetrazolium salt (MTT) solution (5 mg/ml) in 1X PBS and an acidic sodium dodecyl sulphate (SDS) solution (10% in 0.01 M HCl).

This colorimetric assay is based on the cleavage of the tetrazolium ring of yellow MTT by mitochondrial reductase, producing a purple formazan derivative of a quantity that is linearly correlated to the number of viable cells. The water-insoluble formazan crystals are subsequently

dissolved in the acidic SDS solution and the absorbance of the resulting formazan solution is measured by spectrophotometry.

This method was applied to cells grown for a predetermined period on a 96-well or 24-well cell culture plate *(269787, Nunc, Thermo Fisher Scientific* or *83.1836, Sarstedt AG & Co.)*. To the original cell media in the wells, 1.21 mM MTT was added and the plate was incubated in humidified conditions at 37°C and 5% CO_2 for four hours. After incubation, SDS solution was added in a volume equal to that of the original media and the plate was incubated overnight in humidified conditions at 37°C and 5% CO_2. Following incubation, the wells were checked under the microscope for complete crystal solubisation and after 1 minute of shaking, the absorbance was measured at λ = 596 nm with a reference wavelength of λ = 690 nm on a Sunrise Remote microplate scanner *(F039300, Tecan Group Ltd)*.

2.1.9 — Alkaline Phosphatase Activity Assay

Alkaline phosphatase expression is high in undifferentiated ECCs (including 2102Ep and NTera-2 cells) and can be used as a marker for the undifferentiated state of the cell, as this activity reduces upon differentiation.[284] The application of this method to *in vitro* cell cultures provides a simple and semi-quantitative tool to assess the differentiation state of these cells.

Alkaline phosphatase (AP) activity was measured using the Alkaline Phosphatase Detection kit *(SCR004, Merck Millipore)*. This kit contains a Fast Red Violet (FRV) solution (0.8 mg/ml) and naphthol AS-BI phosphate solution (4 mg/ml) in AMPD buffer (2 M). These compounds undergo a reaction catalysed by AP to produce a intensely red-coloured azo-dye. This can be detected by microscopy.

For the assay, cells were seeded on a 96-well cell culture plate *(269787, Nunc, Thermo Fisher Scientific)* and incubated for a specified length of time. Immediately before use, a sufficient measure of staining solution was prepared by combining FRV, Naphthol AS-BI and deionised water (dH_2O) in a microtube at a 2:1:2 ratio respectively. The media were removed from the plate and the cells were fixed a fixing buffer, composed of 10% paraformaldehyde in 1X PBS. After no more than two minutes the fixing buffer was removed and the cells were washed carefully with 100 µl 1X PBS-0.05% Tween (PBST) while ensuring the cells did not detach or dry out. 50 µl staining solution was added to each well and the plate was kept in dark conditions for 15 minutes at room temperature. The staining solution was removed and the cells were carefully washed with 100 µl PBST. A second 100 µl volume of PBST was added and the cells were inspected under the microscope.

2.1.10 — Mycoplasma Contamination Testing

The cells were periodically tested for the presence of mycoplasma contamination. Mycoplasma, a genus of bacteria, are one of the most prolific contaminants of laboratory cell cultures worldwide. This is because mycoplasma are too small to be easily detected by light microscope, the visible signs of their contamination are often absent and they are unaffected by common antibiotics. It is estimated that 11-15% of all laboratory cell cultures in the USA are contaminated with mycoplasma, but his might be higher in other countries or in laboratories where routine mycoplasma testing is not performed.[285] Mycoplasma sometimes have cytopathic effects, but can also cell function in more subtle, less noticeable ways.[285] This introduces undesirable cofactors to an experiment, making mycoplasma detection crucial laboratory practice.

For the cell lines used in this study, mycoplasma testing was performed with the Mycoplasma Detection kit *(11296744001, Roche Applied Science)*. This ELISA-based kit is designed to detect the four most common mycoplasma species (*M. arginini, M. hyorhinis, A. laidlawii, M. orale*). Testing was performed prior to cryogenic cell storage as per manufacturer's protocol.

2.2 — Transfection of Nucleic Acids

Various methods for the introduction of nucleic acid sequences into cells exist, such as chemical transfection, electroporation and viral transduction. All these methods have their benefits and disadvantages and their utilisation is generally based on a variety of factors which can, for example, be polynucleotide type and length specific, target type specific (cells, tissue or whole organisms) or, in the case of cell cultures, cell line specific. In this study modified short RNA polynucleotides and DNA plasmid vectors are transfected into adherent cell cultures. In most cell lines, the transfection method of choice for the transfection of these constructs is by cationic lipid-based transfection agents.

2.2.1 — Methodology

2.2.1.1 — Nucleic Acid Constructs

1. **Pre-miR® miRNA Precursors** *(AM17100, Ambion, Applied Biosystems, Life Technologies)* are synthetic mimics of endogenous miRNA. They consist of double-stranded RNA of which one strand mimics the known mature miRNA. They are chemically modified to resist rapid degradation. Upon transfection of Pre-miRs, the upregulated expression of a specific endogenous mature miRNA is mimicked in vitro and subsequent functional studies or/and target validation studies can be performed.

   For a non-targeting control, Pre-miR® Negative Control #1® (pmNC#1) and Pre-miR® Negative Control #2 can be transfected into cells to assess or cancel out any biological changes directly related to the process of transfection.

   The Pre-miRs for functional upregulation of specific miRNAs were chosen for a number of reasons. The laboratory group had positive experiences using Pre-miRs for cancer research in a variety of cell lines. Pre-miRs have also been frequently and rigorously used in previous experiments involving hESCs and hECCs.[207,265,266,286]
2. **Silencer® Select siRNA** *(4427037, Ambion, Applied Biosystems, Life Technologies)* was used for gene silencing. SiRNA-mediated knock-down of specific gene expression is the dominant method to transiently silence genes in vitro. This method is usually employed to study the phenotypic function of particular genes of interest. More information on the underlying mechanism of gene silencing can be found in Sections 1.8 and 4.1. The Silencer® Select siRNAs succeed the widely used Silencer range (AM16708A, Ambion, Applied Biosystems, Life Technologies). According to the manufacturer Silencer Select siRNAs offer higher specificity and potency than their previous generation due to an improved design algorithm and additional chemical modifications. For a non-targeting

control, the scrambled Silencer® Select Negative Control #1 (ssNC#1) or Silencer® Negative Control #2 can be transfected into the cell line used.

2.2.1.2 — Transfection Agents

In this study, constructs were transfected chemically by cationic lipid-based transfection agents. These agents form structures called liposomes in an aqueous environment. These structures have a positive charge and can incorporate negatively charged nucleotides. The positive charge of the liposome/nucleotide complex, called the transfection complex, aides in the fusion of the complex with the negatively charged cell plasma membrane (PM), releasing its contents into the cytosol. This method of transfection was used as it had successfully been used on the cell lines in this study before. In addition, the method does not require additional equipment besides basic plastic ware.

The transfection agents used for Pre-miR® transfection optimisation were Lipofectamine RNAiMAX *(13778075, Invitrogen, Life Technologies)*, Lipofectamine™ 2000 *(11668019, Invitrogen, Life Technologies)*, siPORT NeoFX *(AM4510, Ambion, Applied Biosystems, Life Technologies)* and DharmaFECT 4 *(T-2004-01, Dharmacon, Thermo Fisher Scientific)*. These were chosen based on their previous successful use by members of the lab and their previous successful use with 2102Ep and NTera-2 cells.[287,288] Furthermore, Lipofectamine RNAiMAX is, as was claimed by Invitrogen, an improved version and less cytotoxic of Lipofectamine 2000, a transfection agent widely used for both cell lines as well as hESCs and other hECCs.[266,289–292]

2.2.1.3 — Transfection Methods

Two methods of transfection are generally used in combination with transfection agents: reverse transfection and forward transfection. In the former method, the transfection agent, transfection vector and cells are mixed in one medium and jointly added to a plate. The latter method involves plating cells in growth medium for a selected period of time previous to transfection to allow them to grow and adhere before exposure to transfection media. The suitability of each method depends on the cell line and on the constructs to be transfected. The advantages and disadvantages are shown in Table 2.1.

Reverse transfection was the initially chosen transfection method as it was a previously proven technique for transfecting siRNA into NTera-2 cells. However, this method has not previously been tested with 2102Ep at the time of this study.[293,294]

As Pre-miRs and siRNAs share a similar structure, chemical make-up and length, a suitable transfection agent should be sufficient for transfecting both constructs.

Table 2.1—The advantages and disadvantages of different transfection methods.

Reverse Transfection		**Forward Transfection**	
Advantages	Disadvantages	Advantages	Disadvantages
Transfection occurs during cell seeding, thus saving time.	Does not provide the opportunity to inspect cell culture conditions and density prior to transfection	Cell culture can be inspected prior to transfection, thus potentially reducing errors.	Cells are seeded and require time to adhere prior to transfection, resulting in a longer protocol
A possible improvement to transfection efficiency due to higher PM exposure to transfection complex.[295]	A possible increase in cytotoxicity and reduced cell viability.	A possibly lower cytotoxicity and higher cell viability.	A possible reduction in cell transfection efficiency. This is cell line-dependent.

2.2.1.4 — Transfection Conditions

The cells are grown in standard hEC growth medium, but during the transfection complex formation and transfection process Opti-MEM® I *(31985-047, Invitrogen, Life Technologies)* is used. This is a serum-free medium and is used because FBS can possibly interfere with the transfection. It is the most widely used medium for nucleotide transfection and has been applied for this purpose to hEC and hESC cells before.[289,296–298] During reverse transfection optimisation OptiMEM I was supplemented with 6% FBS as the cells are maintained in this medium for a longer period of time. During this study an initial assessment was undertaken to study the effect of this medium with 2102Ep and NTera-2 cell lines.

As transfection optimisations and experimental transfections had to be adapted cell culture vessels of different surface sizes, transfections have been tested with 96-, 24- and 6-well cell culture plates and a preliminary or concluding assessment of transfection efficiency was performed.

The transfection agent or the transfection process can be potentially cytotoxicity to the cells, leading to reduced cell viability or proliferation. This can depend on the type and concentration of both the transfection agent and/or the construct used. The effect of the transfection process on cell viability was assessed as detailed in Section 3.3.5.3.

2.2.2 — Methods

2.2.2.1 — Cell Media

During transfection 2102Ep and NTera-2 cells were subjected to a variety of media at different volumes on different plate sizes. These are detailed in Table 2.2. The media during both transfection types additionally contained transfection agent and constructs, both at very low concentrations.

Table 2.2—Different volumes of media used during the transfection experiments.

Cell Culture Plate type	Reverse Transfection	Forward Transfection	Excluding Transfection
96-well	100 µl Opti-MEM® I + 6% FBS	100 µl Opti-MEM® I	100 µl hEC Medium
24-well	600 µl Opti-MEM® I + 6% FBS	600 µl Opti-MEM® I	600 µl hEC Medium
6-well	2.85 ml Opti-MEM® I + 6% FBS	2.85 ml Opti-MEM® I	2.85-3 ml hEC Medium

FBS has been suggested by manufacturers' protocols to interfere with transfection. Thus, the aim was to reduce FBS in the media by combining the serum-free medium Opti-MEM I (*Invitrogen, 31985-047*), containing specific growth factors, with 6% FBS instead of using standard hEC medium containing 10% FBS. Antibiotics (Pen-Strep) are cytotoxic during transfection so they were removed.

2.2.2.2 — Reverse Transfection Optimisation on 96-well Cell Culture Plates

For both cell lines, reverse transfection of Pre-miR® precursors was performed with three transfection agents, Lipofectamine™ RNAiMAX *(13778075, Invitrogen, Life Technologies)*, siPORT™ NeoFX™ *(AM4510, Ambion, Applied Biosystems, Life Technologies)* and DharmaFECT 4 *(T-2004-01, Dharmacon, Thermo Fisher Scientific)*.

Transfection agent volumes were determined in accordance with ranges recommended by the manufacturers. Three volumes were used for each transfection agent; the lowest recommended volume, the highest volume and the mean volume. For ease, the volumes were named Low, High and Medium respectively. Volumes were adapted proportionally to 96-well plate well-size and thus, the cell density used. The total volume of media in each well of a 96-well plate was 100 µl. Table 2.3 shows the volumes of transfection agents used.

Table 2.3—Different transfection agents used at different volumes on 96-well plates.

	Lipofectamine™ RNAiMAX	**siPORT™ NeoFX™**	**DharmaFECT 4**
Low	0.1 µl	0.15 µl	0.05 µl
Medium	0.2 µl	0.675 µl	0.275 µl
High	0.3 µl	1.2 µl	0.5 µl

The stock Pre-miR™ precursors were prepared by diluting 5 nmol Pre-miR™ in 800 µl dH_2O, yielding a concentration of 6.25 µM.

The first step of reverse transfection optimisation was performed by harvesting NTera-2 and 2102Ep cells as described in Section 2.1.5 resuspended in Opti-MEM® I + 7.2% FBS. The cells were counted by the method described in Section 2.1.7 and resuspended in Opti-MEM® I + 7.2% FBS at a specified concentration. Official protocols for the different transfection agents were similar to each other and any differences were incorporated into the protocol of this experiment. All transfection agents expire within 30 minutes after incubation in medium, so strict punctuality was kept to ensure incubation times were not exceeded. 10% overage was maintained where necessary and three biological replicates for each transfection agent dilution were maintained on 96-well plates. Each transfection agent was optimised separately on a separate plate at a separate time.

Pre-miR™ solutions were prepared with Opti-MEM® I (RT°) from the original 6.25 µM stocks. Specified volumes of transfections agents were mixed with Opti-MEM® I (RT°) at dilutions of 10 x final dilutions on the 96-well plate as mentioned in Table 2.3. A transfection master mix was prepared for each transfection agent dilution by combining previously prepared Pre-miR™ solution with previously prepared 10X transfection agent solution and this transfection master mix was incubated at room temperature as shown in Table 2.4.

Table 2.4—Different incubation times for transfection agents.

	Transfection Master Mix Incubation
Lipofectamine™ RNAiMAX	15 minutes
siPORT™ NeoFX™	20 minutes
DharmaFECT 4	20 minutes

To each well of a 96-well cell culture plate, 20 µl transfection master mix was added. This was overlaid with 80 µl of the previously prepared cell suspension. The plate was incubated for 24 hours in a humidified environment at 37°C and 5% CO_2, after which the media were changed to hEC medium and another 24 incubation in the same conditions followed. In addition to the transfected cells, Pre-miR™ controls were also added containing Pre-miR® precursors, but substituting the transfection agent with an equal volume of Opti-MEM® I medium. Furthermore, non-transfected controls were added to the 96-well cell culture plate. These contained untransfected cells cultured in Opti-MEM® I + 6% FBS for 24 hours after which the medium was changed to hEC medium for another 24 hours. Figure 2.1 visualises the entire reverse transfection process.

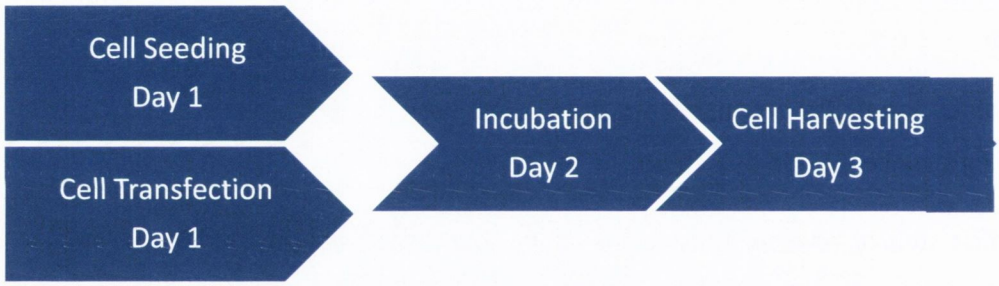


Figure 2.1—The reverse transfection work-flow.

2.2.2.3 — Forward Transfection Optimisation on 96-well Cell Culture Plates

The basic technique of forward transfections are the pre-seeding of cells to the cell culture plate and proceeding with transfection after a period of incubation during which the cells are allowed to adhere and proliferate. Cell health can therefore be roughly asserted with a light microscope before transfection takes place. In addition, as the cells are subjected to transfection conditions for a shorter period of time, forward transfection has the advantage of potentially being less cytotoxic, thus ensuing in more reliable transfection results. Figure 2.2 visualises the entire forward transfection process.

Figure 2.2—The forward transfection work-flow.

For forward transfections, 2102Ep and NTera-2 cells were harvested by the method detailed in Section 2.1.5 resuspended in hEC medium. The cells were counted as per Section 2.1.7, diluted in hEC medium and seeded to wells on a 96-well cell culture plate in 100 µl aliquots. The plates were incubated for 24 hours in humidified conditions at 37°C and 5% $_{CO2}$ to allow the cells to adhere and proliferate. For NTera-2 cells, transfection master mixes for three dilutions of Lipofectamine™ RNAiMAX were prepared in the same manner as described in Section 2.2.2.2. During the incubation of transfection master mixes the cells on the plate prepared 24 hours before were washed with 100 µl 1X PBS (37°C). 80 µl Opti-MEM® I (RT°) was added to each well. To this 20 µl transfection master mix was added. 100 µl hEC medium was added to non-transfected controls instead of Opti-MEM® I and transfection master mix. Compared to reverse transfections (Section 2.2.2.2), the incubation period for the transfection complexes to form was extended to 20 minutes. Additional non-transfected controls were added to the plate, which included the 6-hour incubation step in Opti-MEM® I, excluding any of the transfection components. The plate was incubated for 6 hours in humidified conditions at 37°C after which the media were replaced with hEC medium and the plates were incubated for a specified period.

For forward transfections in 2102Ep cells, transfection master mixes for three dilutions of Lipofectamine™ 2000 were prepared. These dilutions were determined, as with the other transfection agents, within the range of concentrations recommended by the manufacturer, with Low being the lowest, Medium being the average and High being the highest recommended concentration within this range. The volumes used for 96-well plates can be found in Table 2.5.

Table 2.5—Different volumes of Lipofectamine™ 2000 transfection agent for 2102Ep on 96-wells.

	Lipofectamine™ 2000
Low	0.125 µl
Medium	0.25 µl
High	0.375 µl

2102Ep cells were transfected in similar manner to NTera-2 cells. However the manufacturer recommends a five minute pre-incubation of Lipofectamine™ 2000 immediately after dilution in Opti-MEM® I and before adding the transfection constructs. After the pre-incubation, the Pre-miR™ miRNA precursors were added and the complexes were allowed to form during a 15 minute incubation at room temperature. As with NTera-2 forward transfections, hEC and Opti-MEM® I non-transfected controls were included. The cell culture plate was incubated for 6 hours in humidified conditions at 37°C after which the media were replaced with hEC medium and the plates were incubated for a specified period.

2.2.2.4 — Forward Transfections on 96-well Cell Culture Plates

Cells were pre-seeded at 6000 cells/well (approx. $1.8 \cdot 10^4$ cells/cm^2). Only plates which contained cells at an approximate 30-50% density were used for transfections. Transfections of nucleic acid constructs in 2102Ep cells were performed with High (0.375 µl/well) Lipofectamine™ 2000. Transfections in NTera-2 cells were performed with Medium (0.2 µl/well) or High (0.3 µl/well) Lipofectamine™ RNAiMAX, where stated. The forward transfection protocol described in Section 2.2.2.3 was used, substituting Pre-miR-1 for functional nucleic acid constructs used for functional studies. Depending on the application, pmNC#1 or ssNC#1 was used as a negative transfection control. Non-transfected controls containing hEC and vehicle controls, containing only transfection agent, were used to assess any impact of the transfection process on the expression of studied genes/miRNA and on cell viability. All transfections were performed in triplicate.

2.2.2.5 — Forward Transfections on 24-well Cell Culture Plates

Forward transfection methods as described in Section 2.2.2.4 were used for nucleic acid transfection in 2102Ep and NTera-2 cells on 24-well cell culture plates *(83.1836, Sarstedt AG & Co.)*, with a per-well surface area of 2 cm^2. All volumes used for transfections on 96-well plates were scaled up to 24-well plates accordingly. Cells were pre-seeded at approximately 36,000 cells/well, which equates to approximately $1.8 \cdot 10^4$ cells/cm^2. Plates which contained 30%-50% confluent cells were used for transfections. Cells were washed with 1 ml 1X PBS (37°C) prior to transfections. Nucleic acid construct transfections in 2102Ep cells were performed with High (2.21 µl/well) Lipofectamine™ 2000. Transfections in NTera-2 cells were performed with Medium (1.2 µl/well) or High (1.81 µl/well) Lipofectamine™ RNAiMAX, where stated. The forward transfection protocol described in Section 2.2.2.3 was used. Non-transfected controls containing hEC were included. In addition to these, vehicle controls were included, which excluded any constructs from the transfection media. Total volume of media in the wells during transfection was 600 µl. All media were replaced with 600 µl hEC medium (37°C) after transfection. All transfections were performed in triplicate.

2.2.2.6 — Forward Transfections on 6-well Cell Culture Plates

Forward transfection methods as described in Section 2.2.2.4 were used for nucleic acid transfection in 2102Ep and NTera-2 cells on 6-well cell culture plates. Each well of the 6-well cell culture plates used *(83.1839, Sarstedt AG & Co.)* has a surface area of 9.5 cm^2 and all volumes used for 96-well plate transfections were scaled accordingly to keep concentrations constant. Cells were pre-seeded at approximately 171,000 cells/well (approx. 1.8×10^4 cells/cm^2). Plates which contained 30%-50% confluent cells were used for transfections. Cells were washed with 1 ml 1X PBS (37°C) prior to transfections. Nucleic acid construct transfections in 2102Ep cells were performed with High (10.5 µl/well) Lipofectamine™ 2000. Transfections in NTera-2 cells were performed with Medium (5.7 µl/well) or High (8.6 µl/well) Lipofectamine™ RNAiMAX, where stated. The forward transfection protocol described in Section 2.2.2.3 was used. Non-transfected controls containing hEC were included. In addition to these, vehicle controls were included, which excluded any constructs from the transfection media. Total volume of media in the wells during transfection was 2.85 ml. All media were replaced with 3 ml hEC medium (37°C) after transfection. All transfections were performed in triplicate.

2.3 — Total RNA Isolation and Purification

Total RNA isolation included the isolation of all RNA species in the sample. Besides rRNA and mRNA, the RNA isolation methods used were able to also isolate miRNA. In some cases further purification of RNA was necessary to remove impurities which could inhibit down-stream applications. For the same reason, enzymatic DNA digestion was applied to remove residual DNA contamination from all samples.

2.3.1 — Methodology

2.3.1.1 — RNA Isolation with TaqMan® Cells-to-Ct™ Gene Expression Kit

The TaqMan® Gene Expression Cells-to-Ct™ Kit *(4399002 or AM1729, Ambion, Life Technologies)* is a combined solution to cell lysis, DNA digestion, complementary DNA (cDNA) synthesis and qRT-PCR. The kit provides a workflow which excludes a RNA purification step and combines cell lysis and DNA digestion in one procedure. This kit is suitable for downstream applications used in this study for a number of reasons:

1) The protocol significantly economises on time and cost, enabling a high throughput method for gene analysis.
2) The kit furthermore provides an easy solution to obtaining RNA from cells grown on a 96-well cell culture plate as the cells are lysed directly on the plate.
3) According to the manufacturer the kit has higher sensitivity than most other similar products offered as no RNA is lost due to purification and a large volume of cell lysate can be used in the reverse transcription step. This was considered useful to the 96-well plate-based experiments conducted with low cell numbers.

2.3.1.2 — TaqMan® Cells-to-Ct ™ Control Kit

According to the manufacturer's manual the Cells-to-Ct™ kit can be incompatible with certain cell lines due to inhibiting factors in the cell lysate that can hamper downstream applications in the protocol. For this reason, the TaqMan Cells-to-Ct™ Control Kit *(4386995, Ambion, Applied Biosystems, Life Technologies)* is provided to test compatibility. This kit was used with both NTera-2 and 2102Ep cell lines. It comprises of a TaqMan® Gene Expression Assay for ACTB (β-actin), a widely used endogenous control. This control is added to assay to determine whether any reduction in qRT-PCR efficiency is present.

The kit also contains a Xeno RNA control, a synthetic RNA transcript with a non-biological sequence and a TaqMan assay for this RNA. Xeno RNA is added to the samples after cell lysis in

equal quantities. Determining Xeno RNA expression can determine whether the reduction of qRT-PCR efficiency is due to inefficient cell lysis or due to inhibitory factors present in the lysate.

The Cells-to-Ct™ Control Kit tests for inhibitory factors in the cell lysate which can interfere with downstream processes such as reverse transcription or PCR.

The precise protocol for the Cells-to-Ct™ Control kit is described in Section 3.3.2.1.

2.3.1.3 — Kits for RNA Purification

Total RNA purification was performed in experiments for which the Gene Expression Cells-to-Ct™ kit was unsuitable. Two types of RNA purification were used with three different kits, the *mir*Vana™ miRNA Isolation kit *(AM1560, Ambion, Applied Biosystems, Life Technologies)*, the *mir*Vana™ *PARIS kit (AM1556, Ambion, Applied Biosystems, Life Technologies)* and the MagMax™-96 Total RNA Isolation kit *(AM1830, Ambion, Applied Biosystems, Life Technologies)*. Each of these kits were chosen for different applications. Further details are described in Table 2.6.

Table 2.6—RNA purification kits used in this study.

Properties	***miR*Vana™ miRNA Isolation kit**	***miR*Vana™ PARIS kit**	**MagMax™-96 Total RNA Isolation kit**
Isolation Method	Column-based purification with initial acid-phenol:chloroform extraction	Column-based purification with initial acid-phenol:chloroform extraction	Magnetic nucleic acid-affinity binding beads
Application	NTera-2 and 2102Ep cells grown on 24-well and 6-well plates	NTera-2 and 2102Ep cells grown on 6-well plates	2102Ep cells grown on 96-well plates
Additional functions	-	Protein isolation and RNA purification in one kit	DNA digestion step included in kit

DNA contamination can significantly alter the results from downstream experiments such as qRT-PCR or Affymetrix microarray analysis. This is especially true if downstream methods cannot discriminate RNA from DNA. To avoid this, DNA cleanup was performed on all RNA samples used in this study. After initial RNA purification with the *miR*Vana™ miRNA Isolation or *miR*Vana™ PARIS kit, a DNA cleanup was performed with the TURBO DNA-*free*™ kit *(AM1907, Ambion, Applied Biosystems, Life Technologies)*. This was unnecessary for RNA purified with the MagMax™-96 Total RNA Isolation kit, as a DNA digestion step was included in the protocol. The TURBO DNA-*free*™ kit uses an engineered version of DNase I, rendering it more efficient and stable than its wild type counterpart. To stop the DNase reaction, the TURBO DNA-*free*™ kit uses a polysaccharide-base Inactivation Agent to inactivate the enzyme and remove contaminants. This is easier and sacrifices less RNA than approaches involving precipitation or organic extraction. One

consequence from the use of this kit is a lower apparent purity of the RNA as measured by A_{260}/A_{280} and A_{260}/A_{230} ratios. According to the manufacturer, this does not affect downstream experiments such as qRT-PCR or microarray analysis.

2.3.2 — Methods

2.3.2.1 — TaqMan® Cells-to-Ct Gene Expression Kit

The Gene Expression Cells-to-Ct™ kit contains Lysis Solution, DNase I and Stop Solution for the cell combined lysis and DNA digestion stage of the protocol. It furthermore contains 2X RT Buffer and 20X RT Enzyme mix for the cDNA synthesis stage and TaqMan® Gene Expression Master Mix for the qRT-PCR stage. The TaqMan® Gene Expression Master Mix solution is identical to TaqMan® Universal Master Mix, No Amperase® UNG *(4324018 or 4364341, Applied Biosystems, Life Technologies)* and the latter solution is occasionally used as a substitute in this study.

In an aseptic environment, cells cultured for a predetermined period on a 96-well cell culture plate *(269787, Nunc, Thermo Fisher Scientific)* ($<10^5$ cells/well) were washed with 50 µl 1X PBS (4°C) after the media were discarded. Per well 50 µl Lysis Solution (4°C) was prepared with 0.04 U/µl DNase I immediately before the cell lysis step. 50 µl of this solution was used as per manufacturer's protocol with slight modifications; the plate was incubated at room temperature for 12 minutes while manually shaking. This slightly differs from the manufacturer's protocol, which defines an 8 minute incubation, without specifying any agitation of the samples. Downstream qRT-PCR results confirmed the absence of DNA contamination after this modified step, whereas an 8 minute incubation was insufficient to guarantee DNA-free samples. The samples were either used immediately for downstream qRT-PCR or stored at -80°C until further use.

2.3.2.2 — miRVana™ miRNA Isolation kit

The *mir*Vana™ miRNA Isolation kit employs spin column-base RNA purification after an initial RNA isolation step by Phenol:Chloroform extraction. The kit comprises of Lysis/Binding buffer and miRNA homogenate additive for initial lysis. It furthermore contains an acid-phenol:chloroform emulsion for RNA extraction. For RNA binding and purification glass fibre filter cartridges are included with two wash solutions, previously constituted with ethanol and an elution solution (0.1 nM EDTA) for RNA elution.

Adhered cells were washed with 500 µl (24-well plates) or 1 ml (6-well plates) of 1X PBS (0°C) and pelleted cells were washed with 4 ml of 1X PBS (0°C) prior to lysis. Cells were lysed on-plate or in suspension by the addition of 600 µl Lysis/Binding Buffer (RT°C) and overnight storage at -80°C.

Freezing the cells aids their lysis due to the disruption of cell membranes by ice crystals. RNA purification was performed as per manufacturer's protocol, with the addition of rolling the tube horizontally after the addition of wash solutions in order to enhance washing. Also, the ethanol evaporation stage was expanded by leaving the tubes open at room temperature for five minutes. RNA samples were eluted in either 50 µl (24-well plates) or 100 µl (6-well plates) elution solution (95°C) and kept on ice for immediate use or at -80°C for long term storage.

2.3.2.3 — miRVana™ PARIS kit

The *miR*Vana™ PARIS kit is very similar to the *miR*Vana™ miRNA Isolation kit. The main difference is that initial cell lysis is separated in a cell disruption step and denaturing step. Adhered cells on a 6-well plate were washed with 1 ml of 1X PBS (0°C) and pelleted cells were washed with 4 ml of 1X PBS (0°C) prior to lysis. 300 µl Cell Disruption Buffer (0°C) supplemented with 1% v/v protease inhibitor cocktail *(P8340, Sigma-Aldrich Co. LLC)* and 1% v/v phosphatase inhibitor cocktail *(P5726, Sigma-Aldrich Co. LLC)* was added to the samples on ice. This buffer preserves the protein and the protein structure for downstream proteomics and avoids protein denaturing and reduction. Besides the addition of protease and phosphatase inhibitors, it is critical to perform cell disruption on ice as to reduce protease and phosphatase activity. The cell lysates were stored at -80°C overnight to complete disruption. A fraction of cell lysate allocated to downstream protein analysis was stored at -80°C until further use. At room temperature, 2X Denaturing Solution (RT°) was supplemented with 1.5% v/v β-mercaptoethanol *(M-7154, Sigma-Aldrich Co. LLC)* and was added in a volume equal to the fraction of the cell lysate used for RNA purification. RNA purification proceed as per manufacturer's protocol, with the addition of rolling the tube horizontally after the addition of wash solutions in order to enhance washing. Also, the ethanol evaporation stage was expanded by leaving the tubes open at room temperature for five minutes. RNA samples were eluted in 100 µl elution solution (95°) and were kept on ice for immediate use or at -80°C for long term storage.

2.3.2.4 — MagMax™-96 Total RNA Isolation kit

The MagMax™-96 Total RNA Isolation kit was used for RNA purification of 2102Ep cells grown on 96-well cell culture plates. The manufacturer's protocol was substantially modified to improve RNA yield and purity. The kit contains a Lysis/Binding Solution (reconstituted with 100% isopropanol) for initial cell lysis and RNA Binding Beads diluted in Lysis/Binding Enhancer before use are used for nucleotide binding. Two wash solutions (reconstituted with respected 100% isopropanol and 100% ethanol) are used during the wash stages. For the DNA digestion step, TURBO™ DNase with its buffer are included and RNA Rebinding Solution (reconstituted with %100 isopropanol) is used for the rebinding of RNA to the magnetic beads. Finally the Elution Buffer is included for RNA elution. The pure RNA was stored on ice for immediate use or kept at -80°C for long-term storage.

Initial yield and purity of RNA was low so upon advice from the manufacturer and the experience from other laboratories, several modifications were introduced to the manufacturer's protocol:

1. **The samples were frozen at -20°C after addition of Lysis/Binding Solution.** This was considered to improve cell lysis due to the added disruption of cell membranes by ice crystals.
2. **RNA-binding to the beads was improved by extra mixing by pipetting.** This was performed because the beads are heavier than the solutions used and have a tendency to accumulate at the bottom of the well during orbital mixing. Extra mixing resulted in higher RNA yields.
3. **The washing steps and elution steps were improved by facilitating bead dissolution by pipetting after orbital mixing.** The RNA-bound beads remained clumped after orbital mixing and sometimes adhered to the bottom of the plate. Facilitating bead dissolution improved the removal of contaminants from the final purified RNA sample.
4. **Proper DNA digestion was ensured by extending the incubation time from 15 minutes to 20 minutes.** This was performed because trace DNA could still be detected in Minus RT controls during qRT-PCR. Extending the incubation time resolved this problem.
5. **Final ethanol contaminants were kept to a minimum by extending the drying step to four minutes.** Ethanol contaminants dramatically decreased the RNA yield and purity. Additional drying removed any residual ethanol and improved yield and purity.
6. **RNA yield was increased by incubating the samples in 50 µl Elution Solution and incubating them in a 37°C oven for five minutes prior to orbital mixing.** Additional drying, described in modification 5, decreased RNA elution and could be resolved by this additional modification.

2.3.2.5— DNA Removal

For the removal of DNA contamination from RNA samples, the TURBO DNA-*free*™ kit was used for routine or rigorous DNA digestion where applicable. The method was performed as per manufacturer's protocol.

2.4 — RNA Quantification and Qualification

After RNA purification and prior to any down-stream applications, the concentration and quality of all purified RNA in this study was assessed i.e. the RNA isolated with Cells-to-Ct™ Gene Expression kit was excluded from this as these samples did not undergo purification.

2.4.1 — Methodology

2.4.1.1 — NanoDrop 2000

To assess the quantity and quality of RNA in samples, the NanoDrop 2000 *(Thermo Fisher Scientific)* spectrophotometer was used. The NanoDrop 2000 is a micro-volume spectrophotometer with 190 – 840 nm wavelength range, which can measure the quantity and purity of nucleic acid or protein at low sample volumes, producing absorbance curves at high spectral resolution. It uses a patented sample retention system to limit the sample volume needed to 1-2 µl by holding the sample in place by surface retention alone and can measure nucleic acid concentration without the necessity to dilute samples, given the concentrations are < 15 µg/µl. The software automatically calculates, besides sample concentration, the A_{260}/A_{280} and A_{260}/A_{230} ratios for the presence of protein or organic contaminants respectively.

2.4.1.2 — 2100 Bioanalyzer

To assess the quality of RNA used for Affymetrix microarray analysis, 2100 Bioanalyzer *(Agilent Technologies)* was used. This is a micro fluidics-based solution to sizing, quantification and qualification of nucleic acids, proteins and cells. It is a lab-on-chip method of electrophoretic separation and gel analysis, which requires a lower sample volume and provides more qualitative results than traditional gel electrophoresis techniques. A gel image is created from the results which can be visually inspected for RNA degradation products. A secondary, automated method for assessing RNA quality is performed by the 2100 Expert *(Agilent Technologies)* software. From the entire pseudo-electrophoretic trace (electropherogram) the software can calculate an RNA integrity number (RIN). RNA quality can qualitatively and comparatively be assessed with the RIN, which has a value between 1 and 10 with a higher number indicating higher integrity. The RIN is automatically interpreted from key elements of the electropherogram such as 18S and 28S rRNA bands and the presence of RNA degradation products. In this study, the 2100 Bioanalyzer was used for quality control of the RNA destined for Affymetrix microarrays. This has now become the industry standard for RNA evaluation prior to microarray experiments.

2.4.2 — Methods

2.4.2.1 — NanoDrop 2000

For RNA quantity and quality assessment the pedestals of the NanoDrop 2000 were initially washed by mounting 5 µl of nuclease-free and wiping it off. A blank measurement was performed with 1.5 µl nuclease-free water or the solution type similar to that in which the experimental samples were diluted. 1.5 µl sample was measure and the two pedestals were wiped clean in between measurements. The determined quantity, A_{260}/A_{280} and A_{260}/A_{230} ratios were recorded by the software. For both A_{260}/A_{280} and A_{260}/A_{230} ratios the value of 2.1 was considered ideal.

2.4.2.2 — 2100 Bioanalyzer

For RNA integrity analysis on the 2100 Bioanalyzer the RNA Nano 6000 kit *(5067-1511, Agilent Technologies)* is used. This kit contains Nano Gel Matrix and Nano dye concentrate for staining and preparing the RNA. This is loaded on the included RNA Nano chips with the aid of the chip priming station *(5065-4401, Agilent Technologies)* and the IKA vortex mixer *(MS2-S8, IKA-Werke GmbH & Co. KG)*. The kit also contains a reference ladder and a Nano Marker for calibration. Finally two electrode cleaners are included to wash the electrodes with RNaseZap® *(AM9780, Ambion, Applied Biosystems, Life Technologies)* and deionised water before and after use.

The protocol was performed as per manufacturer's instructions. Visual assessment and the RINs were used to determine the integrity of the RNA.

2.5 — RNA Expression Analysis

RNA expression analysis was conducted on RNA samples in this study by relative quantification, i.e. the RNA quantity in an experimental sample relative to a control sample.

2.5.1 — Methodology

2.5.1.1 — TaqMan® Quantitative Real-Time PCR

Ever since the conception of PCR, the uses of this technique have been diversified far beyond its original purpose. QRT-PCR is a modification of PCR which is widely used as an *in vitro* method of accurately quantifying the amount of DNA of a certain sequence and length in a given sample. Although, in purpose, comparable to Northern or Southern blot, the chemistry of qRT-PCR is vastly different. Instead of labelled probes to detect unique DNA sequences, qRT-PCR employs nucleotide primer pairs which can hybridise to unique DNA sequences in a highly specific manner. Amplification of the DNA Section between complementary sequences of the forward primer and reverse primer follows in a manner identical to conventional PCR. However, the addition fluorophores to the PCR process permits a highly sensitive technique to measure DNA amplification after each cycle, i.e. in real-time.

In general, two qRT-PCR approaches are commonly used, which, although based on similar chemistry, employ different mechanisms to achieve accurate quantification of DNA amplification:

1. **Indiscriminate double-stranded DNA-binding dyes**
   With this method conventional, unmodified specific primers are used for PCR amplification of DNA. To the PCR mix, a dsDNA-binding fluorescent or absorbent dye (e.g. SYBR Green I) is added, which indiscriminately binds to any dsDNA present in the reaction. The dye-dsDNA complex can absorb and/or emit light at a specific wavelength and at a respective absorbance and/or intensity proportional to the amount of dsDNA present in the reaction. A post-cycle increase of dsDNA can be accurately tracked during the PCR process , enabling the determination of the concentration of dsDNA in reference to a standard dilution or another dsDNA sequence. The advantages of dsDNA DNA-binding dyes are high sensitivity, the relatively low cost of necessary materials and indiscriminate access to or design of specific primers, however, non-specific binding can deliver false positive results.

2. **Fluorescent probes**

   This technique utilises specific PCR primers and probe, a short nucleic acid with a specific sequence designed to target the amplification region of the template. The probe is chemically modified by the attachment of a fluorescent reporter molecule (e.g. 6-carboxyfluoresein) and matching quencher molecule to each end of the sequence. The proximity of the reporter and quencher prevents fluorescence of the reporter, however when the primer binds to a specific DNA sequence, the subsequent elongation by Taq polymerase cleaves the quencher from the fluorescent reporter, which allows for fluorescence to occur. An accumulation of free fluorescent reporters occurs after each PCR cycle, enabling proportional measurement of starting template in real-time. The advantages of using fluorescent probes are increased reliability as only the specific PCR product is detected. The binding of different types fluorescent reporter molecules enables multiplexing primers in the same reaction. The disadvantages are the running costs and the need to synthesise a different probe for each target sequence.

In this study Applied Biosystems' TaqMan® qRT-PCR was used, which utilises minor groove binder (MGB) fluorescent probe technology. MGB probes have a higher melting temperature allowing shorter probes and higher specificity. In TaqMan® qRT-PCR ROX is used as an internal standard dye with which the overall emission intensity among reactions can be equilibrated.

TaqMan® qRT-PCR consists of two parts:

1. **cDNA Synthesis**

   In this study TaqMan® qRT-PCR is used for the study of mRNA and miRNA expression. As qRT-PCR is functions only with DNA, the RNA molecules have to be converted to DNA. This is performed with an integral part of the TaqMan® qRT-PCR method, cDNA synthesis. This process involves the binding of primers to the template RNA and subsequent reverse transcription by reverse transcriptase (an RNA-dependent DNA polymerase) to create cDNA.

   In this study mRNA and miRNA are reverse transcribed with the High Capacity cDNA Reverse Transcription kit *(4368814, Applied Biosystems, Life Technologies)* and the TaqMan® MicroRNA Reverse Transcription kit *(4366596, Applied Biosystems, Life Technologies)* respectively. Both these kits employ a propriety recombinant Moloney murine leukaemia virus reverse transcriptase, MultiScribe™, for the synthesis of cDNA from the mRNA and miRNA template strands. The two kits differ by their application of primers. For mRNA reverse transcription randomly designed primers which bind and initiate reverse transcription of all mRNA in the sample. For the cDNA synthesis from

miRNAs, conventional random primers are inadequate due to the short length of the miRNA molecule. To address this problem, specific stem-loop primers are used to extend the length of the mature miRNA molecule, while committing only the desired miRNA sequence to the qRT-PCR process, avoiding additional noise in the expression data.

2. **Relative Quantification with Endogenous Controls**

Two methods of quantifying gene expression with qRT-PCR are widely used: absolute quantification and relative quantification. In the first method, gene expression in experimental samples is analysed alongside standard dilutions of known DNA concentration or the analysis of genomic DNA of the specified gene. Absolute quantification can reveal the specific gene copy number per sample or cell. In this study relative quantification is used, which measures the relative expression of specified genes and miRNA to the expression of endogenous control genes. This is useful when comparing gene expression change in a sample with that of a reference sample (usually an untreated control sample or different cell type) without the need of knowing the exact copy number. For comparing these expression changes, the ideal endogenous control gene is one that is invariantly and similarly expressed among all samples in the experiment. Gene expression can be subtly normalised among all samples analysed. The normalised gene expression within each sample (ΔCt) can then be compared among samples (ΔΔCt), resulting in a relative fold-difference of the expression of the studied gene.

Past research has produced a variety of candidate genes which can function as endogenous controls. These genes are often involved in the basic functions of cells due to the invariant necessity of such functions. Examples of endogenous controls and their functions are β-actin (structural support), GAPDH (core metabolism) and 18S rRNA(protein synthesis). Ideally, endogenous controls are of similar length and of similar abundance as the experimental genes as to overcome inaccuracies due to differences in qRT-PCR reaction efficiencies. For miRNA, the aforementioned endogenous controls are often used for normalisation, however due to the low expression of miRNA relative to that of mRNA, and their much shorter length, many studies, including this one, short nucleic acids such as other miRNA or small nuclear RNA (snuRNA), whose stable and invariant expression has been previously verified. This study uses—unless specified differently—GAPDH and miR-16 as endogenous controls for the relative quantification of mRNA and miRNA respectively. The entire TaqMan® miRNA qRT-PCR work-flow and chemical process is summarised in Figure 2.3. The TaqMan® Gene qRT-PCR process is largely similar.

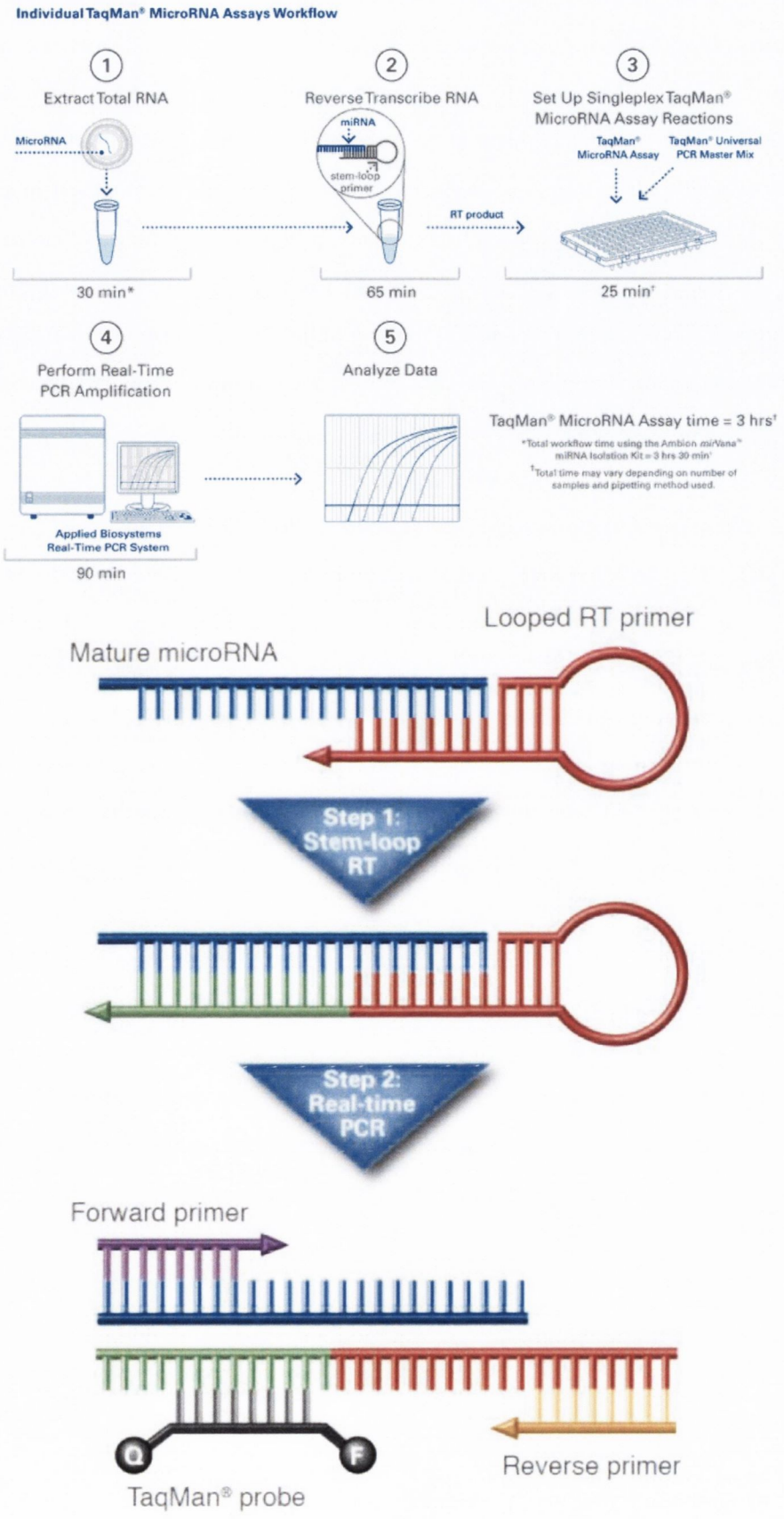


Figure 2.3—The miRNA TaqMan® qRT-PCR work-flow and chemical process. (*ref. Applied Biosystems, Life Technologies)*

2.5.1.2 — TaqMan® Array Human microRNA

As an expansion on TaqMan® technology to its application for high-throughput analysis, Applied Biosystems developed TaqMan® Low Density Arrays (TLDAs) for gene and miRNA expression analysis. TLDAs were later renamed to TaqMan Arrays. These micro-fluidic arrays consist of TaqMan® primer/probe assays for single gene or miRNA targets lyophilised in an array of 384 protected reservoirs (see Figure 2.4). This provides a high-throughput solution to measuring the expression of many different genes or miRNAs in a quantitative way. Each TaqMan® Array card contains 8 loading ports which enable the loading of 48 reservoirs simultaneously with previously prepared cDNA-TaqMan® Master Mix. qRT-PCR can subsequently be conducted in a two-step manner with a total reaction volume of 1 µl. In this study the TaqMan® Array Human microRNA A+B Cards Set v3.0 *(4444913, Applied Biosystems, Life Technologies)* have been used to screen samples for the expression of a total of 754 unique miRNAs. The kit consists of two card types, the TaqMan® Array Human MicroRNA Card A v2.1 and the TaqMan® Array Human MicroRNA Card B v3.0. These cards consist of miRNA primer/probe assays covering the broad majority of the Sanger miRBase v14 miRNA database.[299] Card A v2.1 contains miRNAs with previous demonstrated interest to miRNA research and many are well defined. Extensive optimisation of the primer/probe assays in this pool provides good detection rates. The Card B v3.0 covers miRNAs which have been studied less and are considered to have narrower or lower expression. The limited optimisation of the primer/probe assays on this card produces lower detection rates and increased background signal during qRT-PCR. The TaqMan® Array Human MicroRNA cards were used in conjunction with the Megaplex™ Primer Pools, Human Pools Set v3.0 *(4444750, Applied Biosystems, Life Technologies)*, which contain pools of miRNA RT stem-loop primers covering all miRNAs found on the A and B cards. In addition, miRNA pre-amplification was performed with the included Megaplex™ PreAmp Primers, Human Pool v3.0. Pre-amplification of the miRNA cDNA products are commonly used and highly recommended to enhance the expression signal of miRNAs during qRT-PCR, in particular that of lower expressed miRNAs.

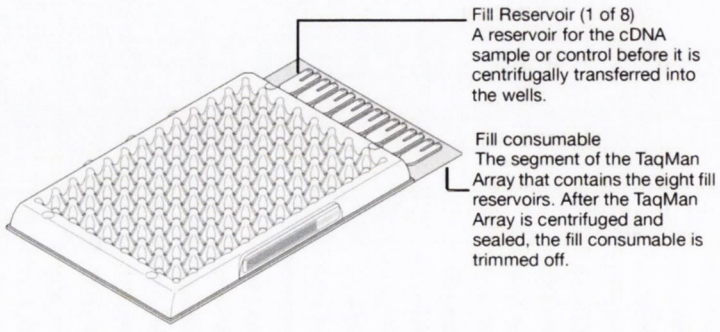


Figure 2.4—TaqMan® Array card. *(ref. Applied Biosystems, Life Technologies)*

2.5.1.3 — Affymetrix Gene Array

Advancing over nearly twenty years, latest generation DNA microarray technology has given researchers the tools to analyse the expression of the entire transcriptome in multiple organisms. Microarrays have become an invaluable method for the study of gene regulation and response by elucidating the differential expression of all genes in given pathways. In addition to gene expression profiling, the use of DNA microarrays has been greatly expanded to many other research domains such as chromatin immunoprecipitation on chip analysis, SNP detection, genome analysis, methylation studies, gene sequencing and the profiling of alternative splicing.

Comparable to Southern blot, but on a high through-put scale, microarray technology is generally based on the principle of labelled DNA hybridisation. A substrate, commonly fabricated from glass or silicon, is spotted with thousands of minute amounts of short DNA strands of known sequence called probes. Sample DNA (reverse transcribed from RNA for gene expression studies) is labelled with fluorescent dyes and hybridised to the microarray chip. A specific hybridisation and wash protocol ensures the correct hybridisation of samples sequences with perfectly complementary probes, while discarding any mismatches. The chip is subsequently scanned on a microarray scanner to produce an image of spots, which fluoresce at an intensity proportional the amount of specific transcripts present in the sample. These fluorescence intensities are converted into values which are background-adjusted and normalised and can be compared to the adjusted intensity values of control samples either hybridised on the same chip with different fluorescent dyes (as is with Illumina gene array technology) or hybridised on separate chips (as is with Affymetrix gene array technology).

For this study, Affymetrix's novel Human Gene 1.0 ST Arrays were used *(901085, Affymetrix Inc.)* to profile the differential expression of mRNA transcripts of known annotation. The Human Gene 1.0 ST Arrays were adapted by Affymetrix from the Human Exon 1.0 ST Arrays and condensed down to include only genes and their isoforms of known annotation as documented by Ensembl (ver. 41.36c), RefSeq and putative complete CDS GenBank transcript databases (as of November 3, 2006). The performance of the Human Gene 1.0 ST Arrays has been validated with the gold-standard GeneChip® Human Genome U133 Plus 2.0 Array, showing very high correlation among results.[300]

The Human Gene 1.0 ST array design is based on the March 2006 (UCSC hg18, NCBI Build 36) human genome sequence assembly and is claimed by the manufacturer to cover >99% of sequences present in the RefSeq database. The Human Gene 1.0 ST Arrays consist of probe sets, pools of individual perfect match probes that target different Sections of the same

oligonucleotide sequence along a whole transcript. On average, four probes per transcript exon and 40 probes per gene are distributed across the chip. In total, 28,132 transcripts with Ensembl support are covered by a total number of 764,885 25-mer probes. Also included on the chip are approximately 17,000 mismatch probes which have been designed to be antigenomic to any known sequences in the human genome. These probes, which are designed to equally include all possible G/C counts in a 25-base oligonucleotide, are used for background signal adjustment after imaging. Also included are 1195 putative exon-level probe sets which target constitutively expressed genes and function as positive controls. Other positive controls on the array are probe sets which target eukaryotic hybridisation controls (bioB, bioC, bioD and creX), 5' and 3' controls and poly-A controls (*lys,phe, thr* and *dap*). These controls function as quality controls for the sample preparation stages and chip hybridisation. 2904 negative control probe sets are also added, which target intronic regions of constitutively expressed genes.

Before chip hybridisation, the samples purified with the method described in Section 2.3.2.2 underwent preparation. For the synthesis of sense strand cDNA from template mRNA Ambion's WT Expression kit *(4411973, Ambion, Applied Biosystems, Life Technologies)* was used, the kit recommend by Affymetrix for this purpose. This kit contains a method based on propriety technology to amplify gene transcripts in the sample and to synthesise single-stranded cDNA containing unnatural dUTP nucleotides. With the GeneChip® WT Terminal Labeling Kit *(900670, Affymetrix Inc.)* the sscDNA was fragmented, labelled and prepared for hybridisation. During the latter stage, the B2 control oligonucleotide and hybridisation controls were added. The former control is used for grid alignment during scanning, while the latter controls are used for quality control of the hybridisation process.

The quality control of the images was performed with Expression Console software *(Affymetrix Inc.)*. The background adjustment, normalisation and analysis of the raw expression data was performed with the Bioconductor 2.9 package for R (CRAN) with the limma plugin.[478] MS Excel 2007 *(Microsoft Corporation)* was used for further analysis of the processed results.

2.5.2 — Methods

2.5.2.1 — TaqMan® Quantitative Real-Time PCR of mRNA

qRT-PCR of mRNA is performed by the two-step TaqMan® method. Purified, DNA-free RNA is diluted to a specified concentration equal among all samples tested in the same experiment. After this cDNA synthesis and quantitative RT-PCR are performed:

1. **cDNA synthesis** is performed with the High Capacity cDNA Reverse Transcription kit *(4368814, Applied Biosystems, Life Technologies)*, which contains 10X RT Buffer, 10X RT Random Primers, 25X dNTP Mix and MultiScribe™ Reverse Transcriptase. All these components are combined according to the manufacturer's protocol to form the RT Master Mix. Preparation of the cDNA synthesis reaction mixes is conducted, as per manufacturer's instructions, in an pyrogen-free environment with pyrogen-free plastic-ware. The total reaction volumes varied according to the amount of cDNA product needed and the individual components of the reaction were proportionally adjusted. The cDNA synthesis reaction is performed in standard 200 µl PCR tubes or on standard 200 µl PCR 96-well plates on a GeneAmp 2400 36-well thermocycler *(Applied Biosystems, Life Technologies)* or a GeneAmp 9600 *(Applied Biosystems, Life Technologies)* 96-well thermocycler. One cDNA synthesis reaction was performed per biological replicate. A negative RT-control was added to each setup by combining all RT components, with the exception of MultiScribe™, replacing it with nuclease-free water. Where deemed necessary, a cDNA positive control was added in the form of RNA with previously validated expression results.
2. **TaqMan® qRT-PCR** of mRNA was conducted with 2X TaqMan® Universal Master Mix, no AmpErase® UNG *(4324018, Applied Biosystems, Life Technologies)*, a mix containing all components necessary for the PCR reaction, except the specific primer/probe mixes and the template cDNA. With a few added changes, qRT-PCR was performed as per manufacturer's protocol in pyrogen-free conditions with pyrogen-free plastic ware on the Prism® 7900HT Sequence Detection System *(Applied Biosystems, Life Technologies)* for 384-well plates or the 7500 Real-Time PCR System *(Applied Biosystems, Life Technologies)* for 96-well plates. TaqMan® qRT-PCR was performed in technical triplicate reactions per sample and a negative no-template control, replacing cDNA sample with nuclease-free water, was added. A few exceptions to the official protocol:

    - 10 µl and 20 µl reaction volumes were used for runs performed in 384-well plates and 96-well plates respectively. This differs from the manufacturer's protocol, which advices 20 µl and 50 µl for these plate sizes. The volume reduction was introduced to save materials and does not significantly affect accuracy.
    - For qRT-PCR reactions conducted on 96-well plates, 5 µl of cDNA product was added to each 20 µl reaction instead of 4 µl. Conversely, the volume of nuclease-free water added was reduced from 5 µl to 4 µl.

The PCR Master Mixes were prepared manually and the MicroAmp® Optical 384-Well Reaction Plates *(4343370, Applied Biosystems, Life Technologies)* or MicroAmp® Optical 96-Well Reaction Plates *(N8010560, Applied Biosystems, Life Technologies)* were either setup manually or automatically with the QIAgility liquid handling robot *(Qiagen Ltd.)*. All plates were sealed with MicroAmp® Optical Adhesive Film *(4311971, Applied Biosystems, Life Technologies)* and briefly centrifuged before run on the aforementioned qRT-PCR thermocyclers.

2.5.2.2 — Cells-to-Ct® Gene Expression

This Section is a continuation of Section 2.3.1.1. Reverse transcription and qRT-PCR were conducted as per manufacturer's protocol. TaqMan® PCR was conducted in similar fashion as described in Section 2.5.2.1.

2.5.2.3 — TaqMan® Quantitative Real-Time PCR of miRNA

As with mRNA, qRT-PCR of miRNA is conducted in a two-step manner:

1. **MiRNA cDNA synthesis** is performed with the TaqMan® MicroRNA Reverse Transcription kit *(4366596, Applied Biosystems, Life Technologies)*, which is similar to the High Capacity cDNA Reverse Transcription kit. The former kit includes RNAse inhibitors to prevent any rapid degradation of low-abundance miRNA in the reaction. No random primers are included as miRNA-specific primers are used. All necessary components were combined according to the manufacturer's protocol (with a few adjustments) to form the RT Master Mix. A total quantity of 10 ng sample RNA was added to each reaction. cDNA synthesis was conducted as per protocol in similar fashion to the method described in Section 2.5.2.1.

   A notable variation to the official protocol was the multiplexing of the RT primers for the experimental miRNA and the endogenous control, miR-16. This was performed to ensure greater reliability of the results by eliminating difference in reaction performance of separate singular reactions. The resultant 3 µl extra volume to the total reaction was compensated by a reduction of nuclease-free water of the same amount. The 10 µl Master Mix volume was plated prior to the addition of the samples RNA. The multiplexing of RT primers was discussed with the manufacturer and was concluded to be possible given the multiplex primers were not complementary and the ratios of components in the reaction were not altered.
2. **TaqMan qRT-PCR** was performed as per protocol in similar manner to Section 2.5.2.1. The qRT-PCR reaction volumes have been reduced to 10 µl and 20 µl for 384-well plates and 96-well plates respectively. When the plate setup was performed on the Qiagility liquid handling robot, the cDNA product was diluted 1:2 to accommodate for the operational margin of the robot. Volumes were adjusted to allow for the addition of cDNA product at the same amount as described in the manufacturer's protocol. TaqMan® qRT-PCR was performed in technical triplicate reactions per sample and a negative no-template control was added.

2.5.2.4 — TaqMan® Array Human MicroRNA

The protocol of cDNA synthesis and qRT-PCR for the TaqMan® Array Human MicroRNA cards is slightly different than that for conventional miRNA qRT-PCR. Besides cDNA synthesis and qRT-PCR, a cDNA preamplification step is included. The protocol is described below:

1. **cDNA synthesis** for the TaqMan® Arrays was performed with the Megaplex® RT Human Pool A and RT Human Pool B in two separate reactions. These are RT primer pools for high-throughput miRNA reverse transcription and require a modified protocol compared with the use of specific RT primers. Reverse transcription is conducted with the TaqMan® MicroRNA Reverse Transcription Kit *(4366596, Applied Biosystems, Life Technologies)* with the 25 nM $MgCl_2$, included with the Megaplex® RT Pools, as an added component. cDNA synthesis is performed as per manufacturer's protocol. The reactions were incubated in standard 200 µl 96-well plates on a GeneAmp 9700 thermocycler *(Applied Biosystems, Life Technologies)* with the Max ramp speed setting.
2. **Preamplification of the cDNA product** was performed with the Megaplex® PreAmp Primers, Human Pool A and Human Pool B. Also included in the protocol was the TaqMan PreAmp Master Mix *(4391128, Applied Biosystems, Life Technologies)*. cDNA synthesis was performed as per manufacturer's instructions and the reactions were incubated in standard 200 µl 8-well strips on a GeneAmp 9700 thermocycler *(Applied Biosystems, Life Technologies)* with the 9600 ramp speed setting.
3. **TaqMan® Arrays** were loaded and analysed as per manufacturer's instructions. The cards were acclimatised to room temperature 30-60 minutes prior to loading. The Pool A and Pool B master mixes for each biological replicate were prepared simultaneously prior to use. While the Pool A card was run (approx. two hours), the Pool B master mix was kept on ice and the B card preparation and loading was timed to be completed upon the availability of the thermocycler. All cards were loaded, centrifuged and sealed with the required equipment detailed in the official manual. The cards were run on the Prism® 7900HT Sequence Detection System *(Applied Biosystems, Life Technologies)*, fitted and calibrated with the TaqMan® Array cycling block.

2.5.2.5 — Affymetrix GeneChip® Human Gene 1.0 ST Arrays

The method for gene expression analysis using the Affymetrix GeneChip® Human Gene 1.0 ST arrays consists of a multistep protocol as described below:

1. **Sample Preparation**

   Sample preparation with the WT Expression kit *(4411973, Ambion, Applied Biosystems, Life Technologies)* was performed as per manufacturer's instructions with some modifications. Prior to preparation the quantity and purity of RNA was measured by the method described in Section 2.4.2.1 and RNA integrity was evaluated with the method described in Section 2.4.2.2. The sample preparation was spread over a 3-day workflow. 300 ng of RNA was spiked with poly-A controls at the correct dilution and was used for first-strand cDNA synthesis.

   First-strand cDNA was synthesised from the template strands. Ambion engineered propriety primer sequences containing a T7 promoter, which selectively bind to mRNA while excluding rRNA present in the sample. Second-strand cDNA was produced by the simultaneous degradation of the original RNA template by RNase H, replacing it with a complementary DNA sequence with the aid of DNA polymerase. From this double-stranded template, cRNA was synthesised by *in vitro* transcription by T7 polymerase for 16 hours and was purified using a magnetic bead-based method similar to what was used in Section 2.3.2.4. Similar modifications were made to the purification protocol:

   - The plate shaking was performed on an orbital shaker *(SO3, Stuart Scientific)*
   - The beads were vigorously vortexed and mixed into the samples by pipetting as an addition to shaking at 500 rpm.
   - Shaking during the washing steps was performed at 800 rpm and during the drying step at 1100 rpm.
   - Dissolution of beads during washing and elution was boosted by pipetting, in addition to shaking.

   Yield and purity of the cRNA was assessed with the NanoDrop 2000 using the method described in Section 2.4.2.1. Second-cycle cDNA was synthesised from the cRNA template, with the inclusion of dUTPs. The original antisense cRNA strand was digested by RNase H and the samples were stored overnight at -20°C. The following day, the cDNA was purified with the same magnetic bead-based method as used previously, however substituting 100% isopropanol with 100% ethanol. Yield and purity were established with the NanoDrop 2000 as described in Section 2.4.2.1. 5.5 µg of cDNA subsequently underwent

fragmentation with the use of the GeneChip® WT Terminal Labeling Kit *(900670, Affymetrix Inc.)* and it's protocol. This kit contains uracil DNA glycosylase and apurinic/apyrimidinic endonuclease 1 which recognise the dUTPs incorporated in the cDNA and cleave the strands at the location of these nucleotides. Terminal labelling of the cDNA fragments was performed with the use of terminal deoxynucleotidyl transferase and biotinylated DNA labelling reagent.

2. **Hybridisation and Scanning**

Prior to hybridisation, the hybridisation cocktail was prepared with volumes specific to the 169 format array and 27 µl of fragmented and labelled cDNA. The 20X Eukaryotic Hybridization Controls were preheated and added to the cocktail last. Each chip was loaded with 80 µl hybridisation cocktail ensuring the presence of an air bubble which was free to move in the chamber. Hybridisation was performed at 45°C for 17 hours in a GeneChip® Hybridization Oven 640 *(Affymetrix Inc.)*.

Staining and washing was performed immediately before scanning with the GeneChip® Hybridization, Wash and Stain Kit *(900720, Affymetrix Inc.)* as per manufacturer's protocol on a primed GeneChip® Fluidics station 450 *(Affymetrix Inc.)*. The chips were registered with the AGCC Fluidics Control Software *(Affymetrix Inc.)*. The spent hybridisation was removed and stored at -80°C and 100 µl Wash Buffer A was injected into the chip. The fluidics station was loaded with appropriate amounts of Stain Cocktail 1, Stain Cocktail 2 and Array Holding Buffer and the chip was stained and washed. After washing and staining the chips were checked for the presence of any air bubbles and they were scanned using a GeneChip® Scanner 3000 7G *(Affymetrix Inc.)*. Figure 2.5 shows a summary of the workflow.

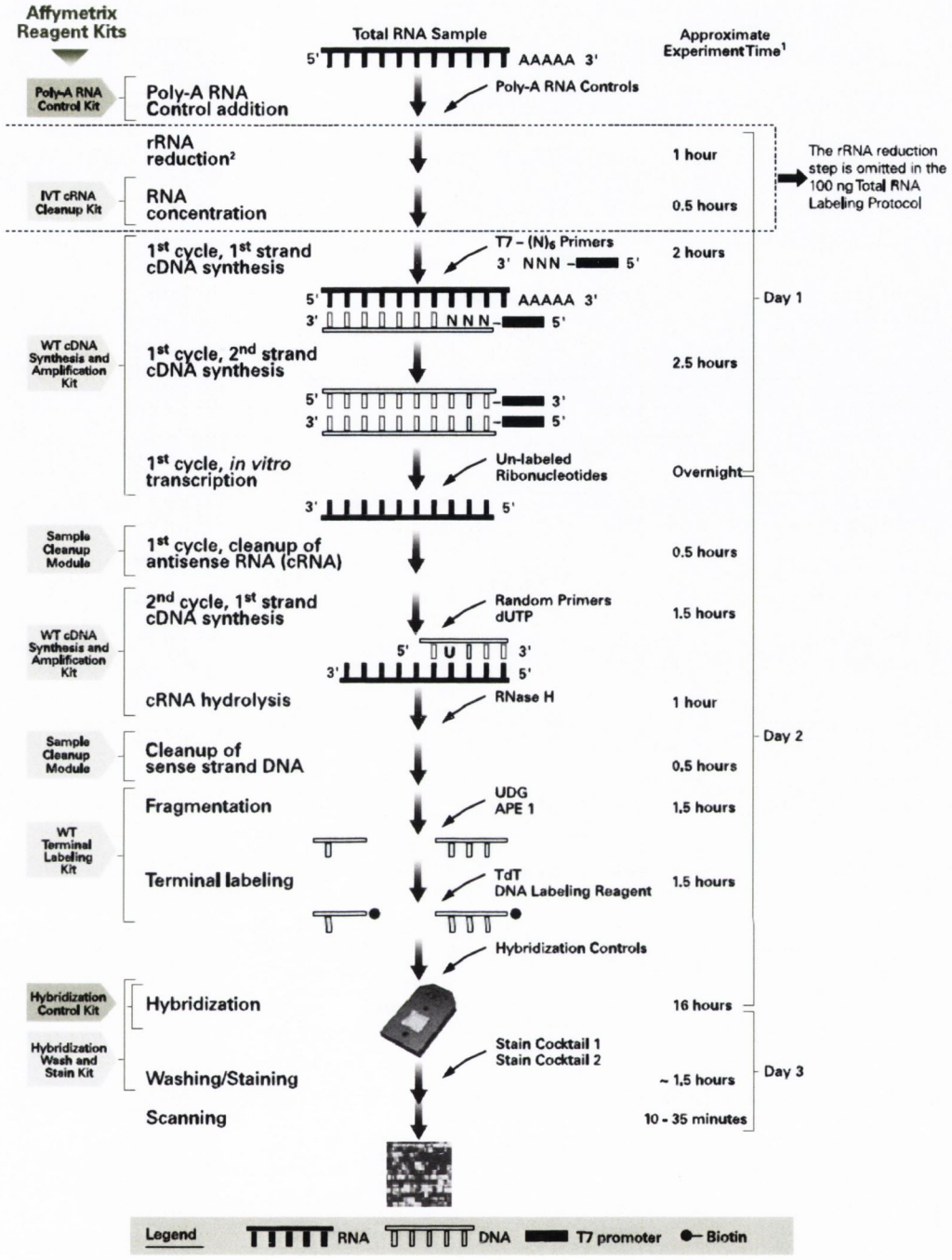


Figure 2.5—General gene array workflow. The general stages of the protocol are: cDNA synthesis, cDNA fragmentation, terminal labelling of the fragments, array hybridisation, washing and scanning.

3. Computational Analysis

The signal intensitiy of ndividual probes was measured by the GeneChip® Operating software and these values were forwarded to the Expression Console® *(Affymetrix Inc.)* software. With this software with which probe sets were summerised and quality control analysis was performed. The original probe intensity values were subsequently summarised again with the same method used for quality control. Probe intensities were also adjusted to background fluorescence and normalised to internal controls.

Subsequently, statistical analysis of the normalised results was performed to produce differential gene expression results between the triplicate non-targeting and experimental controls. The quality control analysis of the gene array experiment, the method of background adjustment, normalisation and differential expression analysis is explained in greater detail in Sections 5.3.2.

2.6 — Protein Expression Analysis

In this study, protein expression was used to determine gene silencing and to validate miRNA targets. The method used to test protein expression is the Western blot method. The whole method for this is described in the next Sections.

2.6.1 — Methodology

2.6.1.1 — Western Blot

Western blot analysis is a widely-used protein immunoblotting method for the semi-quantitative investigation of protein expression and the estimation of their molecular weight *in vitro*. The method uses gel electrophoresis to separate proteins based on their molecular weight or structure which are then transferred to a membrane. The membrane-bound proteins are then probed with specific antibodies. These primary antibodies are subsequently made detectable by the binding of secondary antibodies with specificity to primary antibodies. The secondary antibodies are generally conjugated to a molecule which directly or indirectly allows detection by a plethora of different methods.

In this study, sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) is used for the separation of protein by molecular weight. This is performed after the denaturing of the tertiary protein structures by heat and the denaturing of the secondary structures by the reduction of disulphide bonds. Protein transfer to the membranes is performed with the wet transfer method. After blocking and sequential immunoblotting, the protein bands of specific molecular weight are visualised with enhanced chemiluminescence (ECL). This method utilises an enzyme-linked antibody for the enzymatic catalysis of a chemiluminscent substrate, which, after further oxidation by hydrogen peroxide, produces an excited carbonyl radical that emits light upon decay. This light can be detected and enables a highly sensitive method determining the expression of minute quantities of protein.

2.6.2 — Methods

2.6.2.1 — Protein Extraction

Protein extraction was performed with the Cell Disruption Buffer from the *miRVana* PARIS kit *(AM1556, Ambion, Applied Biosystems, Life Technologies)*. In this study, protein was extracted from cells cultured on 6-well plates. Immediately before use 300 µl/sample Cell Disruption Buffer (0°C) was prepared with 1% v/v protease inhibitor cocktail *(P8340, Sigma-Aldrich Co. LLC)* and phosphatase inhibitor cocktail *(P2745, Sigma-Aldrich Co. LLC)*. The old medium was taken off the cells and they were washed once with 1X PBS (0°C). 300 µl Cell Disruption Buffer was added to each well and the plates were kept on ice for 15 minutes before being frozen at -80°C to enhance cell membrane disruption. The plates were then thawed on ice and the lysate was collected by scraping and transferred to standard 1.5 ml microtubes. The lysates were vigorously pipetted and vortexed to ensure full lysis. Remaining solid cell debris was removed by centrifugation at maximum speed at 4°C for two minutes. The protein-rich supernatant was transferred to new 1.5 ml microtubes and kept on ice for immediate use or stored at -80°C.

Alternatively, where necessary, cells were harvested by means conventional for the cell type, washed with 1X PBS (0°C) and pelleted by centrifugation at 1000·g. Pellets either proceeded immediately to cell lysis or were stored at -80°C. 300 µl Cell Disruption Buffer, spiked with protease/phosphates inhibitor cocktails, was added directly to the fresh or frozen pellets on ice and the lysates were vortexed and pipetted vigorously to ensure full lysis. The lysates were kept on ice for 15 minutes and subsequently frozen at -80°C to enhance cell membrane disruption. After thawing on ice the lysates were centrifuged at maximum speed at 4°C for two minutes to remove cell debris. The supernatant was transferred to new 1.5 ml microtubes and kept on ice for immediate use or stored at -80°C.

2.6.2.2 — Protein Quantification

Protein quantification was performed on all protein samples before further experimentation. Quantification was performed with the BCA Protein Assay kit *(23225, Pierce, Thermo Fisher Scientific)*. This kit uses the property of Cu^{2+} reduction by peptide bonds to Cu^{1+}at 37°C, which then chelates to bicinchoninic acid to produce a purple-coloured product that absorbs light at a wavelength of 562 nm. The kit includes two reagents and bovine serum albumin (BSA) standard ampoules of 2 mg/ml. The protocol is performed as per manufacturer's instructions with modifications to the BSA standard serial dilution in Cell Disruption Buffer to reduce the amount of BSA wasted. On a 96-well plate, 200 µl working reagent was added on 10 µl samples diluted 1:5 with Cell Disruption Buffer and the standards. The plate was left to incubate on in a 37°C oven for

no longer than 30 minutes. A polystyrene foam base was placed under the plate to ensure homogenous heating of all samples and standards. The plate was left to cool at room temperature for 10 minutes before scanning on a Sunrise Remote microplate reader *(F039300, Tecan Group Ltd)* at 562 nm. A linear standard curve was created from the standards with which the protein concentration in the samples was assessed.

2.6.2.3 — SDS-PAGE

In this study, Bio-Rad's Mini-PROTEAN® 3 system *(165-3301/165-3302, Bio-Rad Laboratories)* was used for gel casting and SDS-PAGE. Gel casting is performed with a gel cassette assembly, which enables the casting of mini-gels with a thickness of 0.75 mm sandwiched between two glass plates. Two gel cassette sandwiches can be mounted in the electrophoresis unit, loaded and run together.

The electrophoretic separation of protein was performed with SDS-PAGE. This method uses the protein's electrophoretic mobility to separate proteins, resulting in a protein gradient of decreasing molecular weight. It is important for the proteins to be reduced and heat-denature before running them on the gel, as only peptide strands which have lost their tertiary and secondary structure will organise according to their length.

In this study, SDS-polyacrylamide gels were cast at different specified polyacrylamide concentrations, depending on the molecular weight of the target protein. In general, increasing polyacrylamide concentration will inversely suit proteins of lower molecular weight. In this study most gels were cast at 12% polyacrylamide as most proteins researched lay within the 15-70 kDa range for which this gel provided adequate peptide separation. Each gel consisted of two parts: a smaller 5% stacking gel, containing the sample wells and a larger protein resolving gel. For a 12% gel see the components listed in Table 2.7 were combined.

Table 2.7—Components of a 12% resolving gel used for SDS-PAGE.

12% Resolving gel	Stacking gel
4 ml acrylamide stock (30%)	1.3 ml acrylamide stock (30%)
2.5 ml 1.5 M Tris pH 8.8	1 ml 1 M Tris pH 6.8
3.3 ml deionised water	5.5 ml deionised water
100 µl 10% SDS	80 µl 10% SDS
100 µl 10% ammonium persulphate	80 µl 10% ammonium persulphate
4 µl Tetramethylethylenediamine	8 µl Tetramethylethylenediamine

Tetramethylethylenediamine and ammonium persulphate are responsible for the catalysis of acrylamide polymerisation to polyacrylamide. The anionic detergent, SDS, binds to proteins, causing them to denature and to gain a negative charge. This is crucial for protein fractionation

according to mass as the acquired charge of peptides will be proportional to their length. The gels were precast >12 hours before electrophoresis and the gel cassette sandwiches were wrapped in wet tissue and stored at 4°C overnight to allow full polymerisation.

Accounting for differences in protein concentrations, 10-25 µl samples were diluted 5:6 in 6X reducing protein loading buffer (50 mM Tris-HCl, 0.05% v/v bromophenol blue, 10% w/v SDS, 30% v/v glycerol and 5.5% v/v β-mercaptoethanol) and heated to 95°C for five minutes or otherwise specified to heat-denature the protein. The gels were mounted in the electrophoresis unit and were submerged in 1X Running Buffer (5mM Tris, 50mM Glycine, 0.02% w/v SDS). Although varied, depending on the experiment, on average 15-35 µg protein was loaded per lane in ≤30 µl loading mix. 10 µl Prosieve Color Protein Marker *(50550, Lonza Group Ltd.)* was prewarmed to 37°C and added to a lane. Electrophoresis was performed at 110 V with a power pack *(105, VWR)* until the blue dye reached the bottom of the gel.

2.6.2.4 — Protein Transfer to Membranes

Wet transfer stage followed immediately after SDS-PAGE with the Bio-Rad's Mini Trans-Blot® Cell system *(170-3930/170-3935, Bio-Rad Laboratories)*. Two transfer assemblies (sandwiches constructed with a gel holder cassette holding fibre pads, filter paper, a membrane and the gel) can undergo electrophoretic transfer simultaneously. To construct the blotting sandwich, Immobilon-P™ polyvinylidene diflouride (PVDF) membrane *(Millipore)* was activated for 1 minute in 100% methanol and soaked in ice cold 1X transfer buffer (25mM Tris-HCl pH8.0,0.2M glycine, 20 % methanol). The filter pads and filter paper were soaked in ice cold 1X transfer buffer for <15 minutes prior to assembly. The membrane and gel were sandwiched between two filter papers with two filter pads on the outside and placed in the gel holder cassette after any air bubbles present were pressed out. A diagram of this is shown in Figure 2.6. The assembly was placed between two electrodes in the transfer tank with an ice pack and stirrer to keep the setup cool. Transfer was conducted at 100V for 75 minutes or 45V overnight in a 4°C cold room.

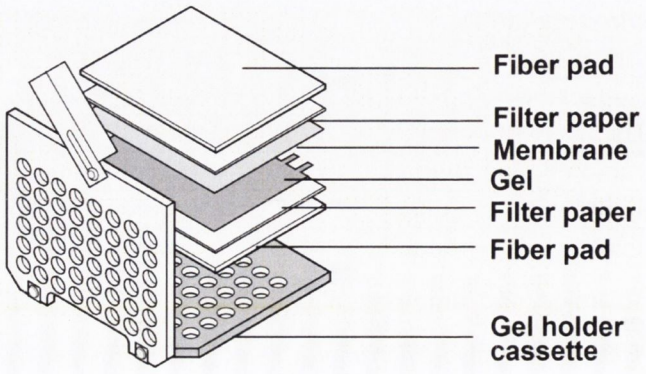


Figure 2.6—SDS-PAGE gel transfer sandwich.

2.6.2.5 — Membrane Blocking and Antibody Binding

After transfer membranes were kept in 1X TBST buffer (10 mM Tris, 0.15 M NaCl, 0.05% Tween-20) at 4°C or they were immediately blocked in blocking buffer (5% w/v fat-free milk *(Marvel, Premier Foods)*in 1X TBST) for >1 hour at room temperature on an orbital shaker. Primary antibodies were diluted in blocking buffer at specified dilutions and replaced the old blocking buffer for an overnight incubation on an orbital shaker in a 4°C cold room. After primary antibody incubation the membrane was washed three times with 1X TBST on an orbital shaker for >5 minutes and a specified dilution of horse-radish peroxidise (HRP)-conjugated secondary antibody in blocking buffer was applied. After >1 hour incubation at room temperature on an orbital shaker or an overnight incubation on an orbital shaker in a 4°C cold room, the membrane was washed three times for >5 minutes in 1X TBST.

2.6.2.6 — Antibody Inhibition

When necessary, inhibition peptides were used to block the primary antibody in a polyclonal antibody solution. The primary antibody to be blocked was diluted in 500 µl 1X PBS (RT°) at a specified dilution and the inhibition peptide was added at a dilution of 5 × [primary antibody]. The antibody-peptide mix was incubated on an orbital shaker at room temperature for two hours and then diluted in blocking buffer before membrane incubation.

2.6.2.7 — Enhanced Chemiluminescence

ECL allows for the visualisation of blotted proteins by the reaction of hydrogen peroxide and luminal to an exited state of 3-aminophthalate, which fluoresces as it decays to lower energy levels. This reaction is catalysed by HRP that, when conjugated to the secondary antibody, provides chemiluminescence localised to the bound primary antibody.

The membranes were incubated for two minutes in ECL solution prepared as described by Haan et al.[301] The membranes were subsequently exposed in an ImageQuant LAS 4000 digital imaging system *(GE Healthcare, General Electric Company)* set for chemiluminescent imaging. The membranes were exposed again to develop a visible-light image of the colour protein ladder. Images were further processed with Adobe Photoshop CS3 *(Adobe Systems Inc.)*.

2.6.2.8 — Membrane Stripping

Membrane stripping was performed by incubating the blotted membranes in Restore Plus Western Blot Stripping Buffer *(46430, Pierce, Thermo Fisher Scientific)* for a defined period of time at room temperature. About 10 ml stripping buffer was used for large membranes and about 6 ml for small membranes. For primary antibodies that were considered to bind their targets weakly, membranes were incubated for 10 minutes in the stripping buffer on an orbital shaker. For strong primary antibodies, membranes were incubated for 15 minutes on an orbital shaker. After stripping, membranes were washed 3 × 20 minutes in 1X TBST after which they were blocked and probed as described in Section 2.6.2.5.

2.7 — Statistical Analysis

2.7.1 — Computational Statistical Analysis

Where relevant, statistical analysis was performed with the Prism ver. 5.0 *(GraphPad Software Inc.)* and SPSS ver. 19.0 *(IBM Corp.)* statistical software packages. These packages provide extensive parametric and non-parametric statistical analysis and both contain a similar set of statistical tests. Prism was used as primary statistical tool, which was replaced by SPSS for tests which were beyond the scope of possibilities of Prism.

For the batch statistical analysis of Affymetrix Gene 1.0 ST microarrays CRAN statistical software *(CRAN)* was used with the Bioconductor 2.9 package. For the batch analysis of TaqMan® Array Human MicroRNA data, DataAssist ver. 3.0 *(Applied Biosystems, Life Technologies)* was used.

CHAPTER THREE

INDUCED UPREGULATION OF MIRNA IN 2102Ep and NTERA-2

3 — Induced Upregulation of miRNA in 2102Ep and NTera-2 Cells

3.1 — Introduction

CSCs are believed to be the causative agent of most cancers.[302] In the heterogeneous model of tumour biology, CSCs are believed to be a subset of cells, which fuel tumorigenesis, maintenance, proliferation and metastasis.[302] Current treatments for cancer, such as surgery, chemotherapy and radiotherapy, take an ubiquitous approach to eliminating malignancy by targeting the whole tumour, attempting to eliminate all tumour cells. This type of treatment is sometimes successful in curing the disease, but success is often related to the type of cancer and its pathological grade (the degree of cell differentiation present).[69,86] In certain types of cancer which are inherently aggressive, relapse is common. Frequently, the cancer returns in a more malignant state and can acquire resistance to the treatment used.[113] Poorly differentiated tumours have a higher CSC-signature and are associated with a poor clinical prognoses.[149,303] The cancer stem cell theory posits the cause of tumour relapse to be a result of an incomplete elimination of primary CSCs.[113] After treatment, a remainder of CSCs can reinitiate tumour growth, frequently leading to a more malignant tumour with more CSCs and an increased resistance due to Darwinian selection.[149,304] This has severe complications for conventional therapy, as it fails to sufficiently discriminate between normal cells, non-proliferating tumour cells and CSCs. Although oncology drugs, due to their proportionally larger investment, are more likely to progress to late phases of clinical trials compared to other types of drugs, and enjoy, on average, shorter regulatory approval times, they have much lower chances of reaching the marketing stage beyond phase III testing (USA data).[305] The latter is most commonly caused by their high cytotoxicity, which due to their failure to focus on the aetiological component of the malignancy, is a severe limitation for medical treatment of cancer patients. 90% of cancer victims die from complications resulting from tumour metastasis.[88] As discussed in Section 1.3, the stem cell phenotype of cancer and its inherent neoplasticity are crucial contributors to the metastatic process. An increasing number of drugs, such as tamoxifen and trastuzumab, can fight metastatic disease, but they often fail to prevent relapse.[306] One proposed reason for this is the diversity of metastases resulting in a lack of drug targeted-effectiveness. Although successful elimination of overt metastatic disease may be clinically established, secondary tumours may reappear from single dormant (quiescent) CSCs or micrometastases, the latter of which probably have CSCs at their root.[306]

The cancer stem cell theory states, that to successfully eliminate cancer completely, all CSCs must be completely eradicated.[307] This requires the development of CSC-targeted remedies. One such method, as discussed in Section 1.4, is to induce CSC-specific differentiation, thus halting their

self-renewal capacity and resulting in cell senescence. To a limited extent, differentiation therapy has already been combined with other chemotherapies in clinical settings of which the use of retinoic acid and other retinoid compounds as chemotherapy and chemoprevention has been most widely studied and used (see Section 1.4 for more details on this).[308] Although this therapy has showed good results, the promising *in vitro* results do not often carry over to clinical results and retinoids can be toxic and possibly teratogenic in clinical doses.[308–311] In addition, the possibility that normal stem cells are affected by this method is quite high, although this has not been documented *in vivo*. Targeted CSC therapy has been proposed as an alternative to current methods (for more details, see Section 1.4). This could potentially eliminate the CSC component of cancer, which would, according to the CSC theory, result in tumour regression and suppression of metastasis.

The therapeutic potential of miRNA-based therapy is discussed in Section 1.11. MiRNAs have an important role regulating the stem cell features of normal stem cells and CSCs, such as self-renewal, proliferation and differentiation. Recently, some of the roles of miRNA in CSC differentiation have been characterised. This showed that various miRNAs with roles in stem cells were deregulated in CSCs, of which the miR-200 and let-7 families were the most common among different CSC types. Some of these miRNAs, such as the tumour suppressor miRNAs, miR-200c and miR-34a, were upregulated in CSCs, which resulted in these cells losing their stem cell phenotype or undergoing apoptosis.[254,269] Utilising the abilities of miRNAs to target multiple genes have been postulated to be of clinical relevance, even when the miRNAs for intended use are not deregulated in the specific cancer or CSC.[246] An ideal treatment would focus on the CSCs, while ignoring normal tissue and normal stem cells. With increased discoveries of CSC markers, some of which may distinguish CSCs from normal stem cells, the possibility of CSC-focussed therapy could become reality. Vectors for *in vivo* delivery for miRNA-based and other RNAi-based therapy are currently being developed. Antibody-conjugated nanoparticles containing anti-c-Myc, anti-MDM2 and anti-VEGF siRNA have been developed that specifically home in on lung metastases of murine melanoma and deliver the siRNA for targeted gene knock-out.[312] This delivery agent can probably also be used with miRNA mimics or antagonists to target a broader spectrum of oncogenes or tumour suppressors.

The search for miRNAs which could potentially regulate crucial CSC processes is ongoing and past findings in cancer and stem cells are currently being directly associated to CSC function. In this chapter, studies were performed to try to find miRNAs with potentially novel roles in CSC and stem cell differentiation. For this two well characterised testicular ECC lines were used: the nullipotent 2102Ep cells and pluripotent NTera-2 cells. Both these cell lines were originally

derived from human testicular teratocarcinomas and later subcloned by Andrews et al.[41,37] 2102Ep cells have the inability to differentiate fully in to the three germ layers of the early embryo, as is characteristic for many ECCs. They show little spontaneous differentiation and are resistant to retinoic acid-induced differentiation.[313] They do however, partially differentiate towards a trophectodermal phenotype when cultured at very low densities.[313] 2102Ep cells have a phenotype that closely resembles that of hESCs , but when they are xenografted into nude mice they form highly undifferentiated teratocarcinomas.[280] NTera-2, a subclone from the Tera-2 cell line, is a pluripotent cell line with the capacity to fully differentiate into cells belonging to the three germ layers when cultured at very low density, but they a have preference to differentiate mostly into CNS neuronal and neuronal precursor cells when exposed to retinoic acid.[314–317] NTera-2 cells form well differentiated teratocarcinomas when xenografted into nude mice.[37] They resemble hESCs quite closely, but fall outside the range of variation of hESC lines.[318,280] In a previous study performed by this laboratory, both cell lines were exposed to retinoic acid for three days and the differential expression of hundreds of miRNAs was measured with qRT-PCR.[274] This study aimed to chart differential miRNA expression during early events of differentiation. As expected, more miRNAs were differentially expressed in NTera-2 cells compared to 2102Ep cells.[274] In the current chapter the data from this study is used to select sets of miRNAs for overexpression in NTera-2 and 2102Ep cell lines, with the aim of inducing differentiation. The criteria used for the selection of these miRNA can be found in Section 3.4.4. Five miRNAs were chosen by these criteria: miR-10a, -99a, -184, -219 and miR-335. The first three miRNAs were upregulated in both NTera-2 and 2102Ep cells upon the incubation of these cells in ATRA for three days. The last two miRNAs were strongly upregulated in NTera-2 cells alone.

One of the aims of this chapter was to predict genes that are potentially targeted by miRNAs of interest—miRNAs which are can regulate hECC differentiation—for post-transcriptional regulation. The prediction of miRNA targets can be performed computationally, with or without support from pre-existing or acquired results. The *in silico* component of miRNA target prediction is under much development and is rapidly changing and adapting to novel findings. Individual miRNAs have been shown to target tens to hundreds of genes in cells.[193] MiRNA can have quite diverse criteria for gene regulation, which make accurate computational target prediction difficult. Various different approaches have been taken to computationally predict miRNA targets. Modern computational target prediction software generally uses a combination bioinformatic algorithms to maximise reliability of target prediction and can rank predicted targets by contextual scores.[319] TargetScan, PicTar and DIANA-microT all use an algorithm which focuses on seed sequence paring, although these prediction tools differ in targeting strategies and hybridisation stringency.[192,320,321] TargetScan is the most frequently used target prediction tool in

this thesis, and its benefits and precise algorithmic execution are described in Section 3.3.9. Other types of miRNA target prediction algorithms are also used in a variety of target prediction software programs. The miRanda algorithm is base on highly complementary sequence alignment between the whole miRNA and its targets. G:U wobbles and mismatches along the hybrid are allowed, but seed pairing has a positive weighted bias towards the contextual score.[319,322] In this thesis, miRanda is used as a secondary target prediction algorithm to complement TargetScan results. Other examples of algorithms are PITA, which combines the free energy gain of miRNA-target hybridisation with the free-energy cost of target mRNA unfolding, thus taking into account the accessibility of the target site for miRNA binding.[197] RNA22 was developed to provides a combination of thermodynamic modelling of the miRNA:mRNA heteroduplex with a search for sequence patterns which are motifs that are significantly shared with mature miRNAs and are likely to function as such.[323] Most target prediction tools search for target sites within the 3'UTR of mRNAs, however PITA and RNA22 can also search outside this region. Also, many prediction algorithms add a variety of weighted factors to the overall calculation of a contextual score. With this, an estimate of target likelihood can be assessed. Examples of such weighted factors are the evolutionary conservation of the miRNA or target site sequences, the presence of multiple target sites and the localised presence of specific bases and regressional ranking of miRNA-target pairs based on previously validated results. Some target prediction algorithms predominantly rely on the regressional analysis of previous results. HOCTAR is a target prediction tool that analyses whole-genome expression data to find statistical anti-correlation between intronic miRNAs (their expression is often correlated with that of their host genes) and potential target genes.[324,325] Other tools, such as mirWIP and starBase, rely on patterns found in validated *in vivo* miRNA-target interactions based on the high-throughput analysis of immunoprecipitated RISC components with which predictive algorithms or statistically qualified target databases can be created.[326,327] No conclusive verdict can be made about which computational miRNA target prediction is most accurate as this field is rapidly advancing and new algorithms are continually being invented and older algorithms often updated. However, much is still unknown about miRNA function and, therefore, computational target prediction in the biological context of the experiment is necessary.

3.2 — Chapter Aims

As no system of oligonucleotide transfection in 2102Ep and NTera-2 cells existed in this laboratory, the first aim of this chapter was to optimise a system for induced *in vitro* overexpression of miRNAs in these cells. This was performed with modified miRNA precursor oligonucleotides as described in Section 3.3.6. The aim was to perform this without substantially affecting cell viability. The second aim of this chapter was to functionally overexpress miRNAs with the intention of inducing observable differentiation in the 2102Ep or NTera-2 cell lines. For any miRNAs which successfully induced the differentiation program, the third aim of this chapter was to find targets that facilitated this process.

3.3 — Materials and Methods

In this Section, methods supplemental to the general methods described in Chapter 2 are provided. These methods are exclusive to this chapter and the referenced Sections in this subchapter supersede the methods described in Chapter 2. If no supplemental methods are given, the general methods are referenced to.

3.3.1 — Cell Culture

With the exception of reverse transfection experiments, 2102Ep and NTera-2 cells were seeded on 96-well plates and 6-well plates as per the method outlined in Section 2.1.4. Cell suspensions were diluted according to the experiment performed and plated on sterile cell culture plates at the desired cell density. Cells were maintained as described in Section 2.1.5. Cells were counted as described in Section 2.1.7.

3.3.2 — Cells-to-Ct™ Gene Expression

The Cells-to-Ct™ Gene Expression kit was used for isolation of RNA and qRT-PCR as described in Section 2.3.2.1 and Section 2.5.2.2.

3.3.2.1 — Cells-to-Ct™ Control Kit

The Cells-to-Ct™ Control kit was used to test the compatibility of the Cells-to-Ct™ Gene Expression kit with the NTera-2 and 2102Ep cell lines. As described in Section 2.3.1.2, this kit contains the non-biological transcript, XenoRNA™ and its associated TaqMan® Gene Expression assay. It furthermore contains a TaqMan® Gene Expression assay for endogenous β-actin. To test compatibility with the Cells-to-Ct™ Gene Expression kit two suspensions (2×10^5 cells/µl) of 2102Ep and NTera-2 cells were prepared in ice cold 1X PBS *(BE17-516F, Lonza)* in microtubes and serially diluted as shown in Table 3.1.

Table 3.1—2102Ep and NTera-2 serial cell dilutions for the Cells-to-Ct™ Control Kit.

	Cells	1X PBS (µl)	Concentration (cells/µl)	Final Cell Count
A	5 µl from 2×10^5 cells/µl stock	45	2×10^4	10^5
B	5 µl from A	45	2000	10^4
C	5 µl from B	45	200	1000
D	5 µl from C	45	20	100
E	5 µl from D	45	2	10

For both cell lines, 5 µl of cell suspension was added to separate 96-well plates in triplicate and kept on ice. Cell lysis was performed by preparing a master mix with a 1:100 ratio of DNase I:Cell Lysis Solution and subsequently adding 50 µl of this master mix to each well of the plate. The

lysate was mixed by pipetting up and down five times and was incubated at room temperature for five minutes. The Stop Solution was spiked with XenoRNA™ to produce a 1:6 ratio of XenoRNA™:Stop Solution and, subsequently, 6 µl of this solution was mixed into each lysate by pipetting five times. The lysates were incubated at RT° for two minutes and subsequently stored at -80°C until further use.

For cDNA synthesis a reverse transcription (RT) master mix was prepared on ice with a per-sample content of 25 µl 2X RT Buffer, 2.5 µl 20X RT Enzyme Mix and 12.5 µl nuclease-free water. For both cell lines, 40 µl aliquots of master mix were distributed in the wells of a single heat-resistant 96-well plate well *(AB-0600, ABgene)* and 10 µl cell lysate was mixed in. To test qRT-PCR sensitivity for any undigested DNA contamination, minus RT control master mixes were added to the plate by substituting the enzyme mix in the RT master mix with an equal volume of water. For both cell lines 10 µl lysate from Sample A was added to the minus RT control wells (see Table 3.1). The plate was sealed with adhesive film *(4306311, Applied Biosystems)* and incubated on a GeneAmp® 9600 thermocycler *(Applied Biosystems, Life Technologies)*, configured with the program found in Table 3.2.

Table 3.2—GeneAmp® 9600 thermocycler program for reverse transcription of sample RNA with the Cells-to-Ct™ Gene Expression kit.

	Stage	Reps	Temperature	Time	Volume
Reverse Transcription	1	1	37°C	60 min	50 µl
RT Inactivation	2	1	95°C	5 min	
Hold	3	1	4°C	∞	

As there was insufficient TaqMan® Gene Expression Master Mix included in the Cells-to-Ct™ Gene Expression kit for the number of 50 µl qRT-PCR reactions required, single replicates were used for the qRT-PCR step. Also, as the samples were already distributed in biological triplicate on the same plates, data reliability was considered sufficient to exclude technical replicates.

With qRT-PCR, β-actin and XenoRNA™ expression was determined. PCR master mixes were prepared containing, per sample, 10 µl 2X TaqMan® Gene Expression Master Mix, 1 µl β-actin or XenoRNA™ TaqMan® Gene Expression assay and 5 µl nuclease-free water. 16 µl aliquots of the master mixes were distributed on a 96-well plate *(N8010560, Applied Biosystems, Life Technologies)* and 4 µl of the respective samples were mixed in. The plate was sealed with optical adhesive film (4311971, Applied Biosystems) and the reaction was run on an Applied Biosystems 7500 Real-Time PCR System configured with the program in Table 3.3.

Table 3.3—GeneAmp® 9600 thermocycler program for qRT-PCR with the Cells-to-Ct™ Gene Expression kit.

	Stage	Reps	Temperature	Time	Volume
Pre-incubation	1	1	50°C	2 min	20 µl
Enzyme Activation	2	1	95°C	10 min	
PCR	3	50	95°C 60°C	15 sec 1 min	

3.3.3 — Pre-miR™ Starter Kit

For transfection optimisation of Pre-miR™ miRNA Precursors, Ambion offers a Pre-miR™ miRNA Precursor Starter Kit *(AM1540, Ambion, Applied Biosystems, Life Technologies)*. This kit contains a Pre-miR-1 precursor, pmNC#1 and a TaqMan® Gene Assay for PTK9. This gene is a validated target of miR-1 and this miRNA has been shown to actively degrade PTK9 at the transcript level, which can be measured with qRT-PCR.[328] As miR-1 expression remains constant in 2102Ep and NTera-2 cells upon exposure to retinoic acid, this miRNA was considered unlikely to alter the cells in a manner which would be undesirable for transfection optimisation. MiR-1 was therefore considered a good candidate for this experiment. Optimising transfection efficiency by measuring PTK9 expression included, besides the efficacy of the transfection of the Pre-miR™ precursor, the efficacy of an increase in miRNA function, which was more important. Besides the Pre-miR™ precursor molecules, the Pre-miR™ Starter kit also contained the transfection agent siPORT NeoFX. This transfection agent was included in the Pre-miR™ transfection optimisation stages of this study.

3.3.3.1 — Baseline PTK9 Expression Analysis

To determine the baseline level of PTK9 expression in unmodified 2102Ep and NTera-2 cells, an initial assessment was performed on GAPDH-normalised PTK9 expression in NTera-2 and 2102Ep cells growing in hEC medium. Three separate 96-well plates were seeded with a triplicate of each cell line at 3.2×10^4 cells/cm^2 in HEC medium at different time points and incubated for 48 hours. In 2102Ep cells, qRT-PCR was performed with purified RNA as described in Section 2.3.2.4 and Section 2.5.2.1. In NTera-2 cells qRT-PCR was performed with the Cells-to-Ct Gene™ Expression kit as described in Section 2.3.2.1 and Section 2.5.2.2.

3.3.4 — Reverse Transfection Optimisation

The first stage of transfection optimisation of Pre-miR™ miRNA precursors was performed with the reverse transfection technique. The underlying principles, benefits and disadvantages of this method is described in more detail in Section 2.2.2.2.

3.3.4.1 — Optimisation of Experimental Media

As described in Section 2.2.2.1, the medium chosen to be used during reverse transfection was Opti-MEM® I supplemented with 6% FBS. The reduced serum content compared to hEC medium was selected so as to minimise any negative effects serum could have on transfection efficiency. To test the compatibility of Opti-MEM® I + 6% FBS medium with the 2102Ep and NTera-2 cell lines, the following experiment was conducted.

On a 96-well plate (*Nunc, 269787*) NTera-2 cells were grown parallel in 100 µl Opti-MEM® I + 6% FBS and hEC medium (DMEM + 10% FBS + 2% Pen-Strep *(Lonza, DE17-603E)* for 48 hours at three different densities (2×10^4, 5×10^4 and 1×10^5 cells/cm^2). Cell viability was subsequently measured using the MTT method, as described in Section 2.1.8. In total, two experiments were performed in biological triplicates and the results were combined.

3.3.4.2 — Optimisation of Reverse Transfections

The optimisation of reverse transfections of Pre-miR® precursors in 2102Ep and NTera-2 cells were performed on 96-well cell culture plates using the methods described in Section 2.2.2.2.

3.3.5 — Forward Transfection Optimisation

With the forward transfection technique an attempt was made to increase transfection efficiency while reducing cytotoxicity caused by the transfection process. The benefits and disadvantages of forward transfection are described in Section 2.2.1.3. For the optimisation of forward transfection, some individual components were optimised first as described in Section 3.3.5.1 to Section 3.3.5.3.

3.3.5.1 — Optimisation of Experimental Media

The media used for forward transfection was serum-free Opti-MEM® I medium (RT°). Transfection in 2102Ep and NTera-2 cells was planned to be conducted for 6 hours with this medium before being replaced by hEC medium (37°C). Before transfection optimisation was conducted, cell viability in the serum-free conditions of Opti-MEM® I was tested.

2102Ep and NTera-2 cells were harvested as detailed in Section 2.1.5 and resuspended in hEC medium. The cells were counted as per Section 2.1.7 and diluted to a concentration of 6×10^4 cells/ml. From this stock dilution 100 µl was seeded in triplicate to wells on a 96-well cell culture plate and they were incubated for 24 hours in humidified conditions at 37°C and 5% CO_2. After incubation, the cells were washed with 100 µl 1X PBS (37°C) and 100 µl Opti-MEM® I (RT°) was

added to six wells. To another six wells hEC medium (37°C) was added to function as negative controls. The plates were incubated for 6 hours in humidified conditions at 37°C and 5% CO_2, after which the old media were replaced with hEC medium (37°C). The cells were incubated for 48 hours, with a medium refresh after 24 hours. Cell viability was assessed as described in Section 2.1.8.

3.3.5.2 — Optimisation of Forward Transfections

The optimisation of forward transfections of Pre-miR® precursors in 2102Ep and NTera-2 cells was performed on 96-well cell culture plates using the methods described in Section 2.2.2.3.

3.3.5.3 — Cell Viability

The impact of the forward transfection process on cell viability of 2102Ep and NTera-2 cells was assessed by the transfection of ssNC#1 using the forward transfection method. For 2102Ep cells the Low (0.125 µl/well), Medium (0.25 µl/well) and High (0.375 µl/well) Lipofectamine™ 2000 volumes were applied. This was similar for NTera-2 cells which used Low (0.1 µl/well), Medium (0.2 µl/well) and High (0.3 µl/well) volumes of Lipofectamine™ RNAiMAX. Transfection was performed in triplicate using 50 nM pmNC#1 for both cell lines. For transfection, the method described in Section 2.2.2.3 was applied. Negative controls were prepared by preseeding cells in Opti-MEM® I and incubating them for 6 hours on the same plate. As with the experimental cultures, the medium was replaced with hEC medium (37°C) and the cells were incubated for another 48 hours, with a medium refresh after 24 hours. Cell viability was assessed for all cells on the plate by the protocol described in Section 2.1.8.

3.3.6 — Functional Pre-miR™ Transfections

The transfection of experimentally functional Pre-miR™ precursors and pmNC#1 controls in 2102Ep and NTera-2 cells grown on 96-well cell culture plates were performed as described in Section 2.2.2.4. The method of transfection of these molecules into cells grown on 6-well plates is described in Section 2.2.2.5. For both plate sizes, cells were transfected with 50 nM Pre-miR™ precursor. Details on the Pre-miR™ precursors used in this chapter are found in Table 3.4.

For gene and protein expression analysis, these constructs were transfected into 2102Ep and NTera-2 cells pre-seeded on 6-well cell culture plates. The plate layout is shown in Figure 3.1. On 96-well cell culture plates, all transfections were performed simultaneously on the same plate. MiR-10a, miR-99a and miR-184 were transfected with High Lipofectamine™ 2000 and with Medium Lipofectamine™ RNAiMAX™ in NTera-2 cells. After transfection the cells were left to grow for three days with new hEC every 24 hours. After 48 hours following transfection 2102Ep

cells were split 1:2 using the method described in Section 2.1.5. 2102Ep cells double very rapidly (doubling time: 21 hours), so a split was performed to aid phenotypic visualisation of cells, which would otherwise be hindered by their high confluence on the third day of incubation. During the post-transfection incubation period images were taken at different magnifications with a light microscope.

The transfection experiments with miR-219 and miR-335 were performed once the transfections with miR-10a, miR-99a and miR-184 were completed. For the transfection of miR-219 and miR-335 certain modifications were introduced as to improve the contrast and reliability of the results. Initial miRNA expression results from transfections with these miRNA showed poor miRNA upregulation in NTera-2 cells (data not shown). The decision was taken to adjust the Lipofectamine™ RNAiMAX™ volume upwards to High volumes (8.6 µl/well) for NTera-2 cells. This was because concurrent transfections with other construct types performed by colleagues in the laboratory showed improved transfection efficiencies at these Lipofectamine™ RNAiMAX concentrations for NTera-2 cells grown on 6-well plates. After transfection the cells were left to grow for 6 days with new hEC every 24 hours. The cell culture plate layout is shown in Figure 3.1 and details about the Pre-miR™ precursors used are listed in Table 3.4. Three days post-transfection, the cells were split 1:2 using the method described in Section 2.1.5. One split was cultured overnight for photographic imaging with a light microscope.

Table 3.4—Product type and sequence of Pre-miR™ precursors for functional miRNA overexpression.

miRNA	Product ID	Mature Sequence
hsa-miR-10a	PM10787	UACCCUGUAGAUCCGAAUUUGUG
hsa-miR-99a	PM10719	AACCCGUAGAUCCGAUCUUGUG
hsa-miR-184	PM10207	UGGACGGAGAACUGAUAAGGGU
hsa-miR-219	PM10664	UGAUUGUCCAAACGCAAUUCU
hsa-miR-335	PM10063	UCAAGAGCAAUAACGAAAAAUGU

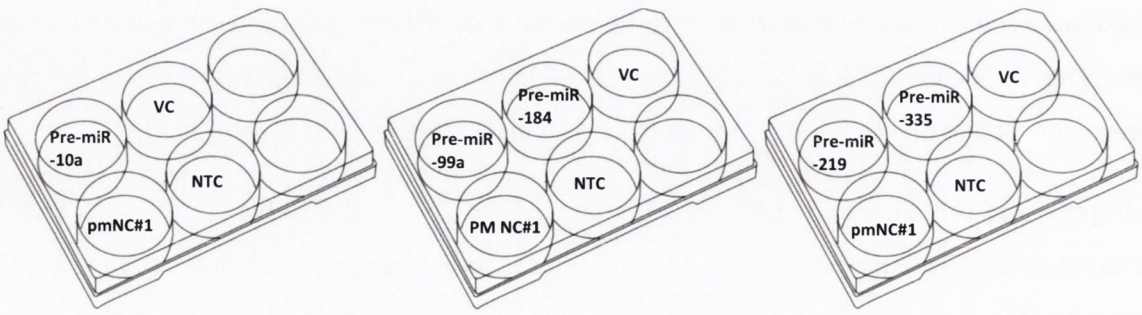


Figure 3.1—Cell culture plate layout for functional transfections of Pre-miR™ precursors. Pre-miR-10a and Pre-miR-99a/Pre-miR-184 were transfected on different plates. Pre-miR-219 and Pre-miR-335 were transfected on the same plate. Each plate included negative controls: pmNC#1, vehicle control (VC) and non-transfected control (NTC). This image represents a single biological replicate.

3.3.7 — Analysis of miRNA and Gene Expression

After transfection and incubation, the cells were lysed and RNA purification was performed in accordance with the protocols for 6-well plates detailed in Section 2.3.2.3. DNA digestion was performed prior to qRT-PCR using the method described in Section 2.3.2.5. RNA was quantified with the NanoDrop™ 2000 as described in Section 2.4.2.1 and the RNA concentrations of the samples were normalised by dilution with dH_2O. MiRNA expression was measured using the protocols found in Section 2.5.2.3. The miRNA expression assays used for TaqMan® PCR are listed in

Table 3.5. All expression results were normalised to miR-16 and relative quantitation was applied with the $2^{-\Delta\Delta Ct}$ method. Previous studies performed by this laboratory extensively demonstrated a constant expression of miR-16 in differentiated and undifferentiated 2102Ep and NTera-2 cells, making it a good normalisation control for miRNA qRT-PCR.

Table 3.5—Product codes of TaqMan® primer/probe assays used for miRNA expression analysis.

miRNA	AB Assay ID
hsa-miR-10a	000387
hsa-miR-99a	000435
hsa-miR-184	000485
hsa-miR-219	000522
hsa-miR-335	000546
hsa-miR-16	000391

The expression of key pluripotency markers and differentiation markers was measured using the original RNA used for measuring miRNA expression. RNA was diluted to the required amount for all samples and qRT-PCR was performed as described in Section 2.5.2.1. Gene expression results were normalised to GAPDH and the relative expression was calculated using the $2^{-\Delta\Delta Ct}$ method. The TaqMan® gene expression assays used are listed in Table 3.6.

Table 3.6—Product IDs of TaqMan® primer/probe assays used for pluripotency and differentiation marker analysis.

Gene	Assay ID
OCT4	Hs01654807_s1
SOX2	Hs00602736_s1
NANOG	Hs02387400_g1
ENO3	Hs00266551_m1
NCAM1	Hs00169851_m1
GATA6	Hs00232018_m1
GAPDH	4331182

3.3.8 — Protein Expression Analysis

Protein concentration was assessed with the BCA method described in Section 2.6.2.2. Sample replicates were run on separate 12% gels using SDS-PAGE as described in Section 2.6.2.3. A lane with 10 µl Prosieve colour marker *(50550, Lonza Ltd.)* was included. A 75 minute transfer to PVDF membrane was performed as described in Section 2.6.2.4. Subsequent blocking and overnight probing in primary antibody was performed as described in Section 2.6.2.5. Primary and secondary antibodies used at specific dilutions can be found in Table 3.7. The anti-mouse secondary antibody from Jackson Immunoresearch (115-165-062) was only used for binding to the anti-GAPDH primary antibody. For all other mouse antibodies, the antibody from Santa Cruz (sc-2314) was used. After a 3 x 5 minute wash in 1X TBST buffer the membranes were exposed as described in Section 2.6.2.7. All membranes were stripped as and reprobed as described in Section 2.6.2.8.

Methods of computational miRNA target prediction have been constantly evolving from basic mismatch seed-target based pairing to a plethora of different methods, which often maintain different criteria with the aim to most closely reflect the *in vivo* processes of miRNA-enabled post-transcriptional regulation. Numerous software tools have been released to apply these methods to miRNA target prediction *in silico*. Examples of the criteria and considerations made in the methods which these tools apply to are miRNA-mRNA hybridisation energies, phylogenetic conservation of miRNA target sites, seed-sequence or whole-miRNA incomplete complementarity and mRNA region-accessibility. Although accuracy differs among the various tools available, no ideal software package exists which predicts *in vivo* miRNA-target binding with perfect accuracy.

Table 3.7—Primary antibodies for Western blot.

Gene	Predicted MW	1st Antibody ID	Clone Type	1st AB Host	Dilution	2nd Antibody ID	Dilution
SOX2	34.3 kDa	ab75485	Mono	Mouse	1:500	sc-2314	1:2000
OCT4	38.6 kDa	sc-5279	Poly	Rabbit	1:500	sc-2313	1:5000
GAPDH	36.1 kDa	ab8245	Mono	Mouse	1:25,000	115-165-062	1:2000

3.3.9 — Computational miRNA Target Prediction

In this study, the TargetScan v5.1 software tool was chosen for the prediction of miRNA targets.[191,192,329] TargetScan is one of the most frequently used target prediction tools and has been frequently praised for its relatively high accuracy (estimated >50-60% of predicted genes are targeted *in vivo*).[330] This online tool performs thermodynamics-based imperfect seed-sequence matching to the 3' UTRs of mRNA sequences found in the NCBI RefSeq database, accessing 17821

different genes in this database. TargetScan requires an exact match of ≥7 bases in the seed sequence.[192,319] When a gene has multiple mRNA isomers, TargetScan selects the RefSeq accession with the longest 3' UTR. The software selects mature miRNA sequences from the Sanger miRBase miRNA database, however some sequences are modified to better reflect recent findings. The seed sequence is defined by the software as the region between the 2-8 nucleotides from the 5' end of the miRNA and a minimum 6 nucleotide match of the seed sequence with the 3' URT region is required. The software filters and ranks the target mRNAs based on the evolutionary conservation of miRNA target sites within their 3' UTR. The software checks for target-site conservation in the genomes of 23 different organisms and categorises the target in a ranking according to phylogenetic proximity to humans. In this study, targets of poor conservation (i.e. not conserved in vertebrates) were excluded. TargetScanHuman also has some disadvantages i.e. the seed-sequence flanking regions and mRNA target site accessibility are not included in the algorithm.

3.4 — Results

In this Section, the results of the transfection optimisations with Pre-miR™ miRNA precursors and the functional Pre-miR™ transfections are shown. All transfection optimisations and functional transfections were performed with 2102Ep and NTera-2 cells.

3.4.1 — Pre-optimisation Control Experiments

3.4.1.1 — Optimisation of the Cells-to-Ct™ Gene Expression Kit

The Cells-to-Ct™ Gene Expression kit was used for cell lysis and qRT-PCR during transfection optimisation experiments for reasons described in Section 2.3.1.2. At the time of this study, no published research demonstrated the use of the Cells-to Ct™ Gene Expression kit with the 2102Ep and NTera-2 cell lines, thus the compatibility of this kit for this study had to be established. This experiment was performed as described in Section 3.3.2.1.

The Ct values of the β-actin marker and the XenoRNA™ marker, obtained during qRT-PCR, were plotted against the cell numbers given in Table 3.1. The curves for NTera-2 and 2102Ep cells can be found in Figure 3.2. As the Ct values increase or decrease by the exponent of two, the ideal Ct values for the β-actin marker would logarithmically and inversely correlate to cell numbers. As an equal amount of this synthetic nucleic acid was added to each sample, the ideal Ct values for XenoRNA™ should be exactly equal for each cell number. If inhibitory factors are present in the lysates, and consequently in the qRT-PCR reactions, we would expect the β-actin marker Ct values for the higher cell numbers to lose their logarithmic consistency with the other results and gradually skew to lower exponents. The XenoRNA™ marker Ct values would lose the linear consistency at the higher cell numbers and would skew to higher values. To determine the "non-linearity" of the results after log-transformation of the cell number values, the significance of deviation from a linear regression trend—this is the only expected trend for both β-actin and XenoRNA™ values against log-transformed cell numbers—was calculated using the replicates test lack-of-fit method as described.[331,332]

Results show that for NTera-2 cells, β-actin expression obtained with qRT-PCR is linearly proportional to the cell numbers used in each sample tested as the results do not significantly deviate from a linear regression model ($P=0.5528$). The picture is different for results obtained from 2102Ep cells, in which β-actin expression steadies at the higher cell numbers. This non-linear change digresses significantly from a linear regression ($P<0.0001$).

XenoRNA™ expression trend shows no significant change for increasing NTera-2 cell numbers ($P=0.3903$), however for 2102Ep cells, XenoRNA™ expression progressively reduces when tested in the presence of higher cell numbers ($P<0.0001$).

These results clearly showed that NTera-2 cells were compatible with the Cells-to-Ct™ Gene Expression kit at the cell numbers used, whereas 2102Ep cells were clearly incompatible with this kit. For 2102Ep cells, the MagMax™-96 Total RNA Isolation kit was therefore used for RNA isolation and the High Capacity cDNA Reverse Transcription kit was used for cDNA synthesis (see Section 2.3.2.4. and Section 2.5.2.2).

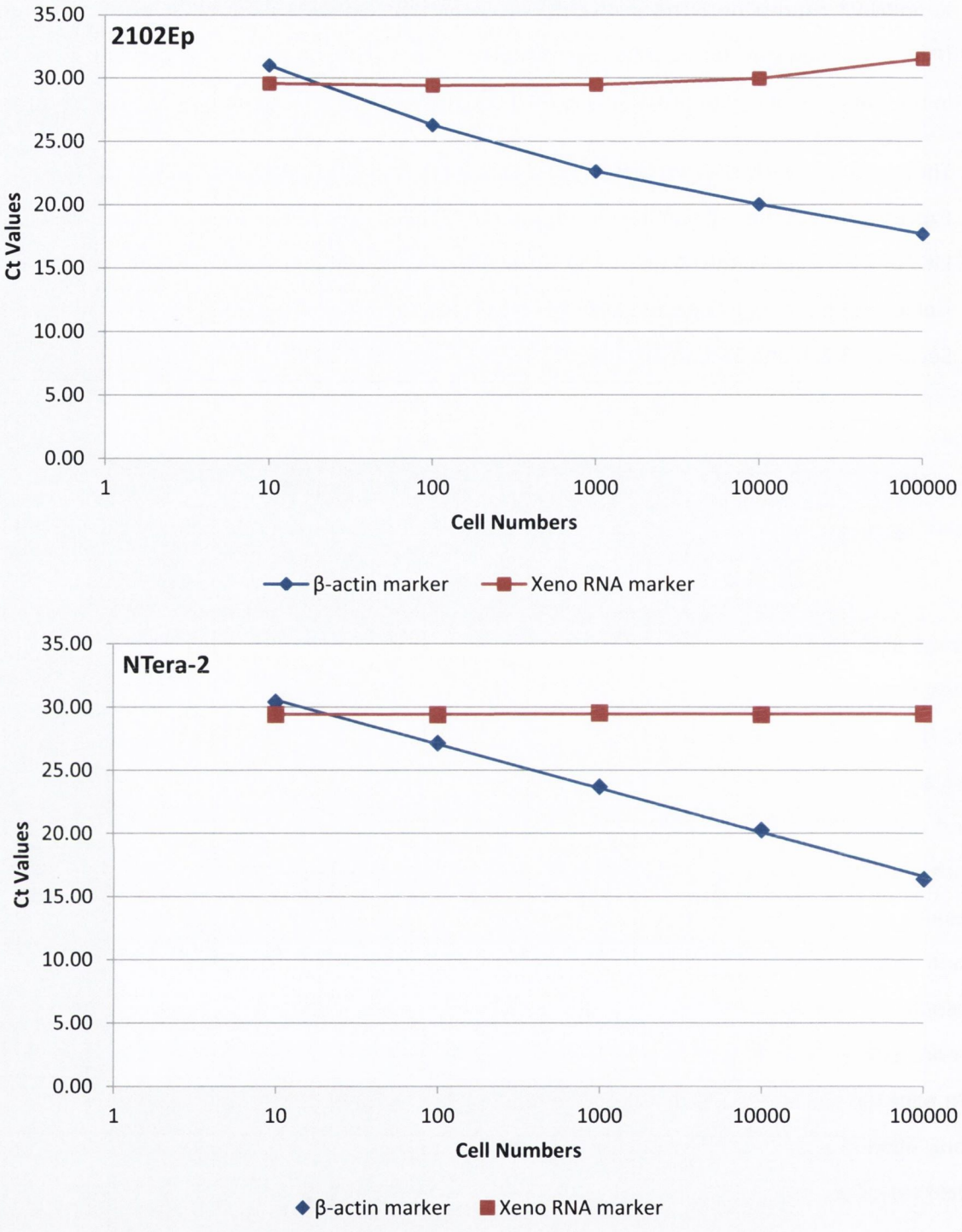


Figure 3.2—Ct expression values of β-actin and XenoRNA™ versus cell numbers in 2102Ep and NTera-2 cells. Both cell lines were serial diluted and mixed with XenoRNA™. A linear trend in expression of β-actin and XenoRNA™ indicates cell line compatibility with the Cells-to-Ct™ Gene Expression kit. In 2102Ep cells, the β-actin and XenoRNA™ marker values both deviate significantly from a linear trend (both $P < 0.0001$) whereas in NTera-2 cells these values do not significantly deviate from this trend ($P = 0.5528$ and $P = 0.3903$ respectively). Biological triplicates were used for all dilutions. Values are mean Ct expression. A lack-of-fit test was used to determine significant deviations from linearity.

3.4.1.2 — Determining Baseline PTK9 Expression in Both Cell Lines

The Pre-miR™ Starter kit utilises the expression of the gene, PTK9, to assess successful transfection of Pre-miR-1. Before Pre-miR™ transfection optimisations could be commenced, baseline PTK9 expression had to be determined in 2102Ep and NTera-2 cell lines. This assessment was performed in unmodified 2102Ep and NTera-2 cells as described in Section 3.3.3.1. The raw expression values expressed in Ct were normalised with the endogenous control, GAPDH, resulting in δCt values as displayed in Figure 3.3.

In both cell lines PTK9 is sufficiently expressed with raw Ct values of approximately 26 in 2102Ep and 27 in NTera-2 cells. This is within the ideal 15-35 Ct range of reliable expression detection with qRT-PCR. Also, little variance between replicates in both cell lines exists. From these results the conclusion was drawn that PTK9 was a successful control for transfection optimisation.

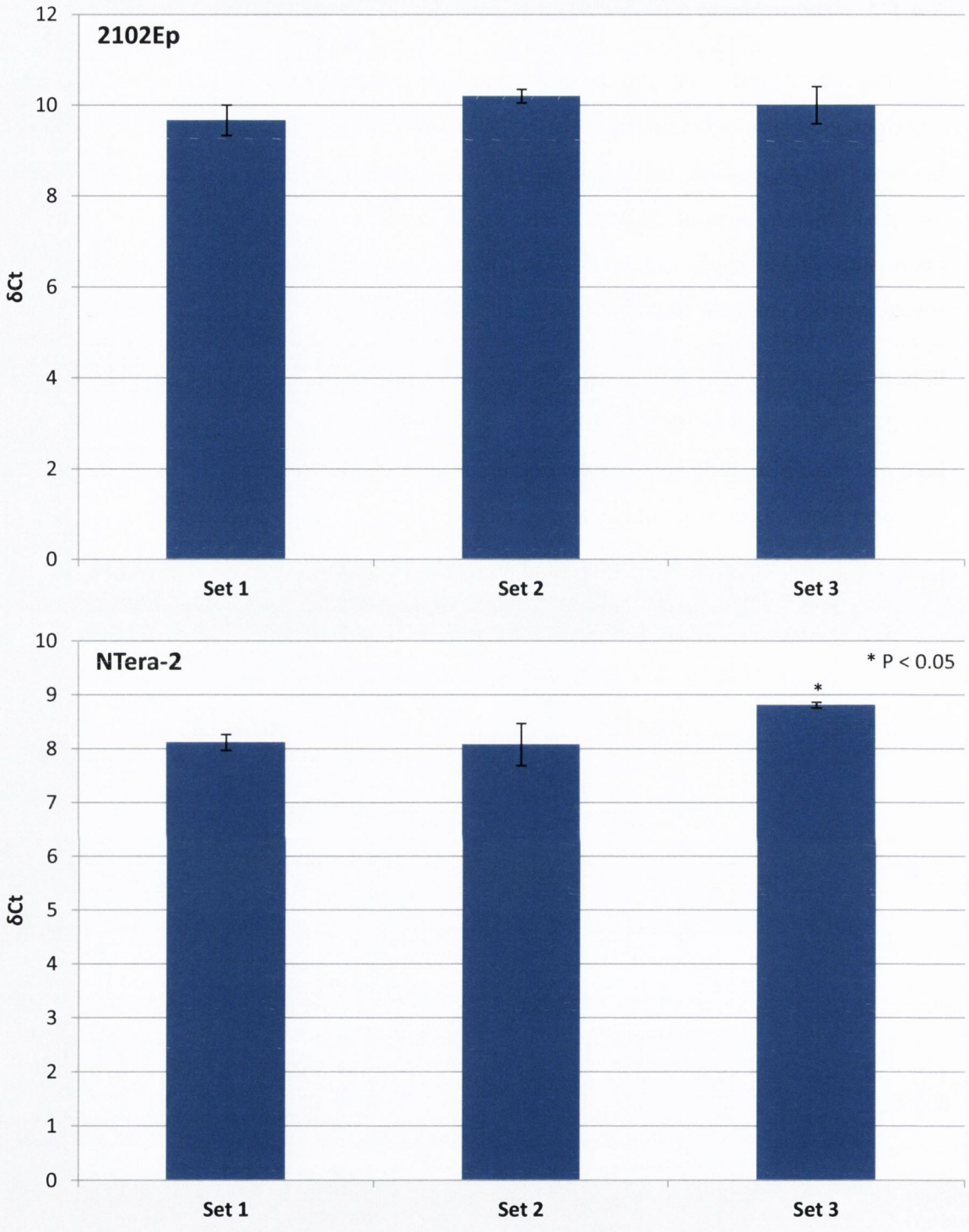


Figure 3.3—The comparison of PTK9 expression normalised to GAPDH expression in 2102Ep and NTera-2 cells. Normalised PTK9 expression was measured in both cell lines during their normal, proliferative state. In 2102Ep cells, the average δCt of PTK9 among all sets and replicates is 9.95. No significant difference in relative expression between sets ($P = 0.2014$). In NTera-2 cells, the average δCt of PTK9 among all sets and replicates is 8.37. Here, ANOVA shows significant difference among sets ($P = 0.0174$) and Tukey's post-hoc tests show that Set 3 differs significantly from the other sets in NTera-2 cells. Values are normalised mean expression ± SD of biological triplicates.

3.4.2 — Reverse Transfection Optimisation of Pre-miR™ miRNA Precursors

As no protocol for the transfection of nucleic acids into 2102Ep cells was established in the laboratory, the transfection of Pre-miR™ miRNA Precursors had to be optimised for this cell line. The first stage of optimisation was performed by reverse transfection, as this was considered to improve transfection efficiencies at the cell densities required for these cell lines, while resulting in significant time savings on the transfection protocol. GAPDH was used as a normalising control for PTK9 expression analysis of all transfections as this gene was highly expressed in both cell lines. It did not change during retinoic acid exposure to these cell lines and was previously used as an endogenous control in transfection studies in these cell lines.[285,326–328]

3.4.2.1 — Optimisation of Experimental Media

As described in Section 2.2.1.4, the reverse transfection process was performed in Opti-MEM® I medium supplemented with 6% FBS. To assess the effects of this medium on medium-term cell growth 2102Ep and NTera-2 cells were both subjected to Opti-MEM® I + 6% FBS medium and hEC medium as described in Section 3.3.4.1.

The results of cell viability are displayed as raw absorbance data to describe relative change and are shown in Figure 3.4.

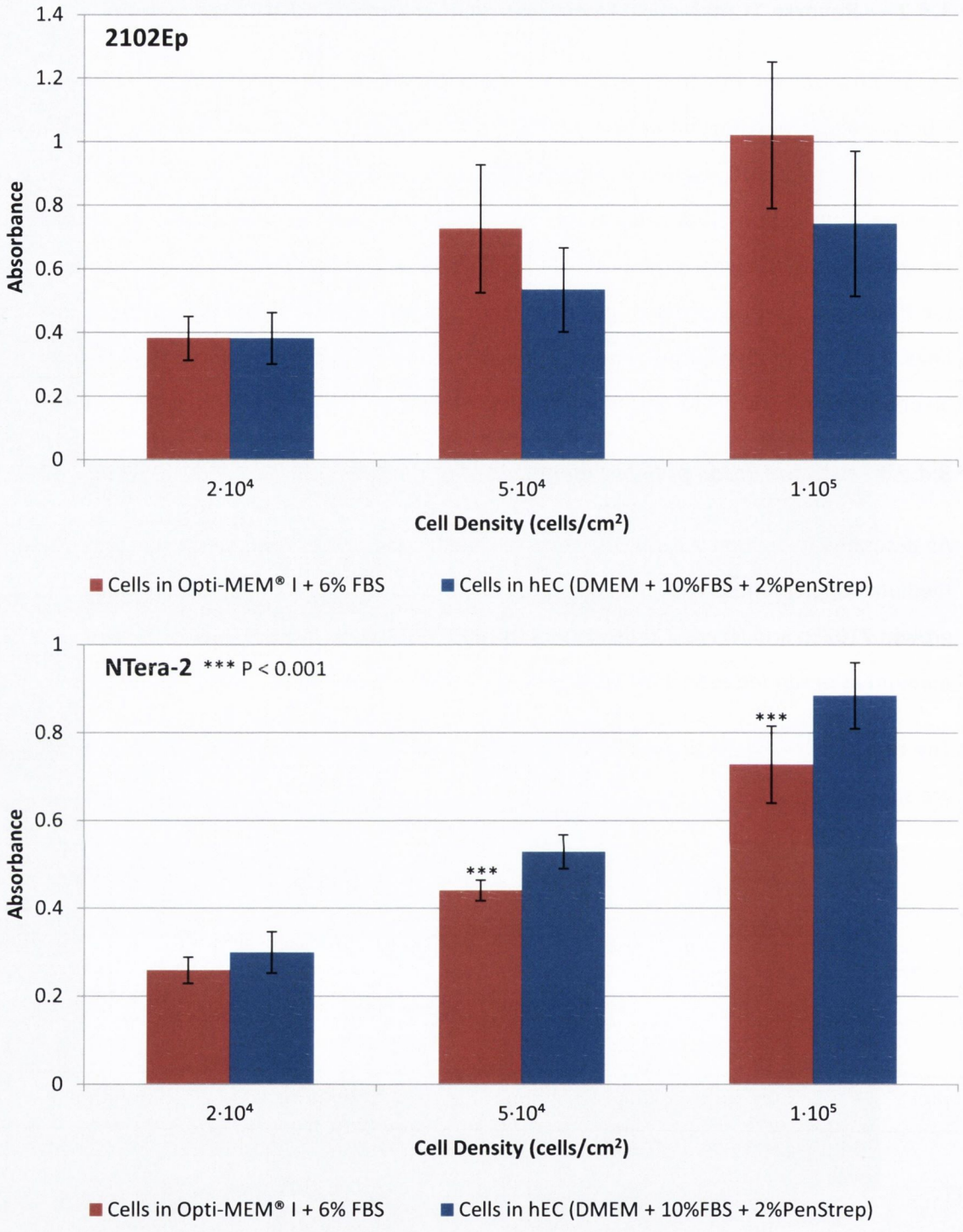


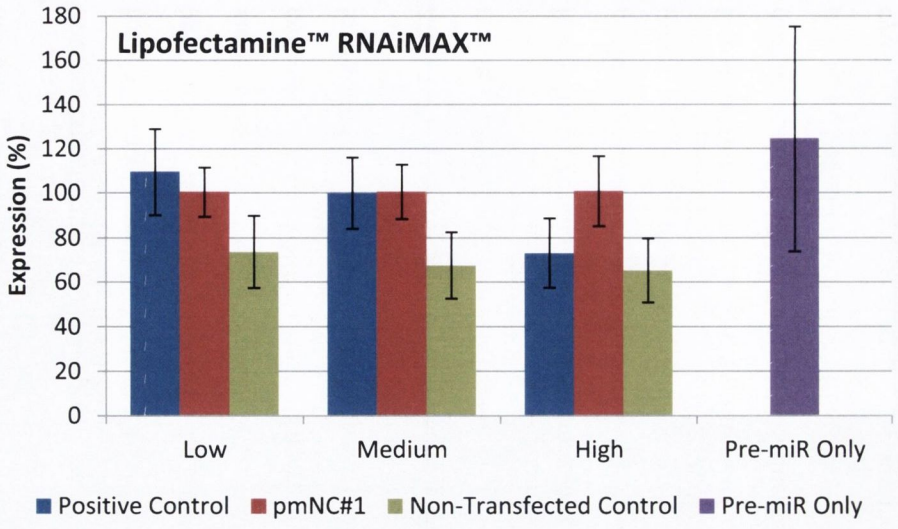
Figure 3.4—Cell viability of 2102Ep and NTera-2 cells grown in Opti-MEM® I and hEC media. Cells were grown in the mentioned media at three different cell densities for 48 hours. Among 2102Ep cells no significant differences in cell viability exist between Opti-MEM® I + 6% FBS and hEC medium at any density ($P > 0.05$). Among NTera-2 cells no significant difference exists in cell viability between cells grown in Opti-MEM® I + 6% FBS and hEC medium at 2×10^4 cells/cm^2 densities ($P = 0.1062$). At 5×10^4 and 1×10^5 cells/cm^2 densities viability of cells in Opti-MEM® I + 6% FBS is significantly lower ($P = 0.0007$ and $P = 0.0076$ respectively). Values are mean absorbance ± SD of six biological replicates. A t-test was used to compare sample means.

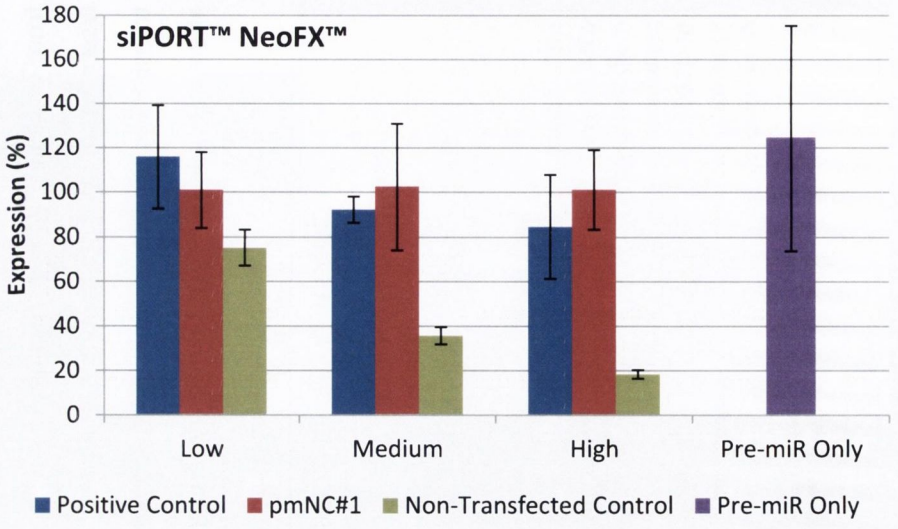
3.4.2.2 — Optimisation of Reverse Transfection in 2102Ep Cells

Reverse transfections in 2102Ep cells were performed with Lipofectamine™ RNAiMAX™, siPORT™ NeoFX™ and DharmaFECT 4 at three different concentrations using the method described in Section 3.3.4.2. As mentioned in this Section, the concentrations of transfection agent chosen were based on the recommended concentration range by the manufacturers. The "High" concentrations are the highest concentration recommended, the "Low" concentrations are the lowest and the "Medium" are the mean concentrations within the range.

After a 24 hour transfection and another 24 hour incubation of the 2102Ep cells, PTK9 expression results were obtained and displayed in Figure 3.5.

From these results the conclusion could be drawn that none of these transfection agents performed adequately to downregulate PTK9 expression by reverse transfection with 50 nM Pre-miR-1. Some results showed a limited knock-down, however this was insignificant compared to the non-transfected controls. Unusually, PTK9 expression seemed to decrease progressively relative to the negative controls with increasing siPORT™ NeoFX™ concentration. This is probably due to potential undesirable effects the transfection agent has on GAPDH expression due to the cytotoxic nature of the transfection agent, which can affect GAPDH expression according to the manufacturer. This claim is supported by the decrease of raw GAPDH expression (inferred directly from Ct values) with increasing transfection agent concentration. These effects were seen to a lesser extent in Lipofectamine™ RNAiMAX™ and DharmaFECT 4 transfected cells. Due to the poor transfection efficiencies and highly probable cytotoxic effects of the transfection method, an alternative approach was taken as described in Section 3.3.5.





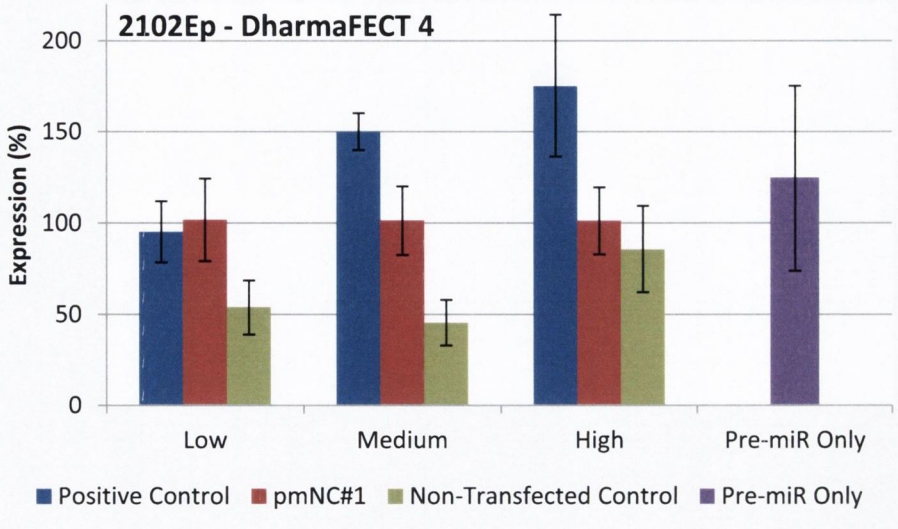


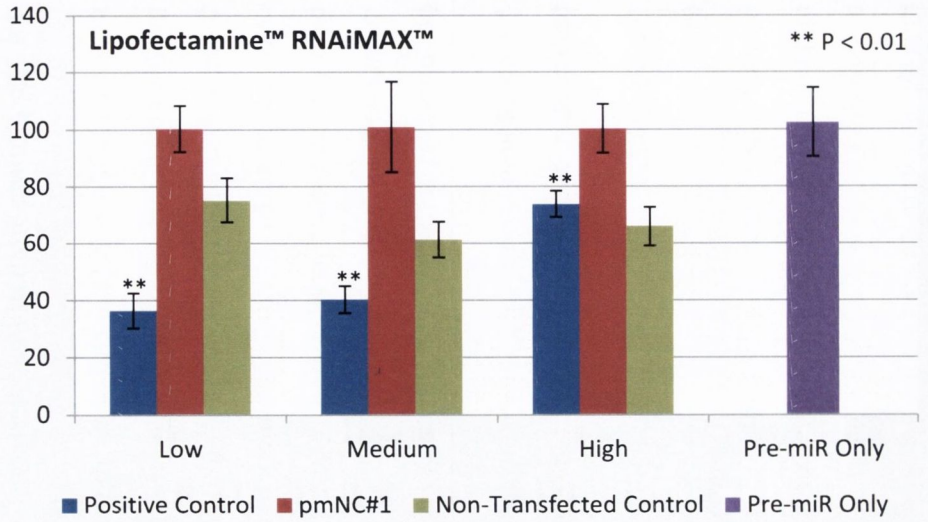
Figure 3.5—Expression of PTK9 in 2102Ep cells reverse transfected with miR-1 three three different volumes of Lipofectamine™ RNAiMAX™, siPORT™ NeoFX™ and DharmaFECT 4. When compared to the pmNC#1-transfected cells, no significant PTK9 suppression by miR-1 could be detected in any of the transfected cells with any of the transfection agents tested. However, in some cases PTK9 expression was lower in the non-transfected controls than the pmNC#1-transfected cells. For cells transfected in Lipofectamine™ RNAiMAX™ expression values were significantly lower in Non-Transfected Controls compared to pmNC#1 cells transfected in Medium Lipofectamine RNAiMAX ($P = 0.0413$) and High Lipofectamine RNAiMAX ($P = 0.0446$) conditions. For cells transfected with siPORT™ NeoFX™ expression values were significantly lower in Non-Transfected Controls compared to Negative Controls transfected in Medium siPORT NeoFX ($P = 0.0156$) and High Lipofectamine RNAiMAX ($P = 0.0013$) conditions. Expression values were significantly lower in Non-Transfected Controls compared to pmNC#1 cells transfected in Low DharmaFECT 4 ($P = 0.0372$) and High DharmaFECT 4 ($P = 0.0125$) conditions. The expression of PTK9 in the pmNC#1 samples was set at 100%. All experiments were performed in biological triplicate. Values indicate the mean of relative expression ± SD. Sample means were compared with a t-test.

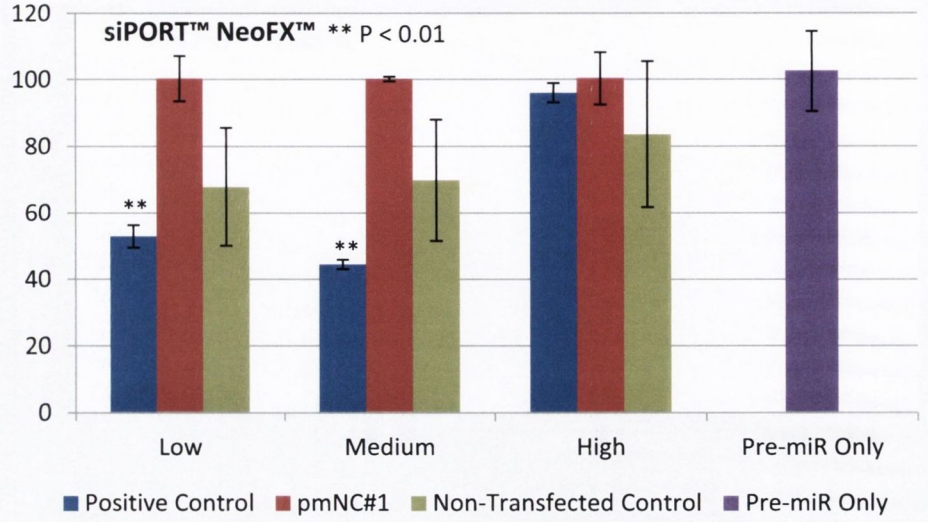
3.4.2.3- Optimisation of Reverse Transfection in NTera-2 Cells

As with 2102Ep cells, the three transfection agents used for reverse transfection of Pre-miR™ miRNA Precursors were Lipofectamine™ RNAiMAX, siPORT™ NeoFX™ and DharmaFECT 4. Of these, only DharmaFECT 4 was previously used in a reverse transfection experiment with NTera-2 cells by a research group at the time, however this was not performed with the same construct type (i.e. Pre-miR™ miRNA precursors).[287] The same transfection agent concentrations were maintained as were used with 2102Ep reverse transfections. NTera-2 transfections were performed as described in Section 3.3.4.2.

After a 24 hour transfection and another 24 hour incubation of the NTera-2 cells, PTK9 expression results were obtained and displayed in Figure 3.6.

The transfection of Pre-miR-1 with all three transfection agents resulted in the downregulation of PTK9. The best knock-downs reported were with Low DharmaFECT 4 (0.05 µl/well) and Low and Medium Lipofectamine™ RNAiMAX (0.1 and 0.2 µl/well respectively). At these conditions the remaining PTK9 expression was approximately 34-40%. As this was still considered insufficient to ensure the induction of a phenotypic effect during the functional overexpression of experimental miRNAs an alternative approach was taken for transfection optimisation. This choice was also made because , as with 2102Ep cells, reduced GAPDH expression in the Lipofectamine™ RNAiMAX and siPORT™ NeoFX™ transfected cells indicated potential cytotoxic side-effects of the reverse transfection process.





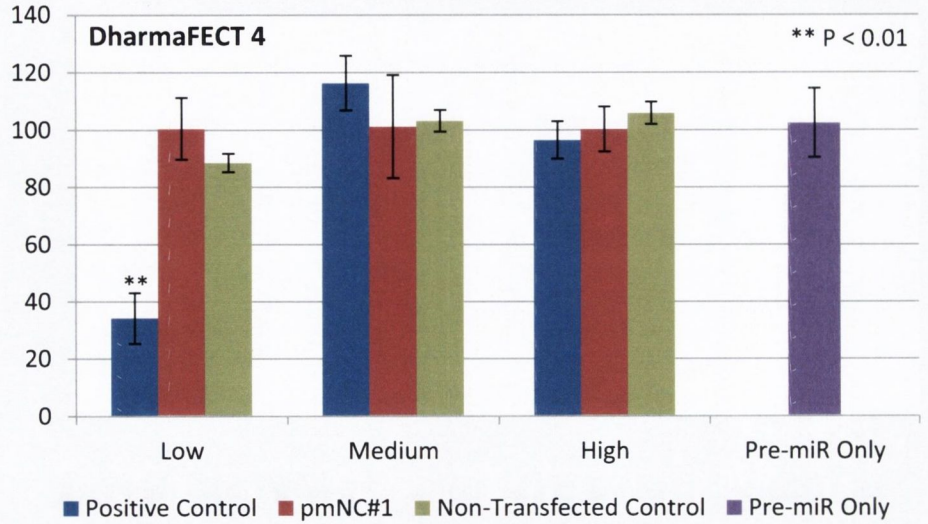


Figure 3.6—Percentage expression of PTK9 in NTera-2 cells reverse transfected with miR-1 using three different volumes of Lipofectamine™ RNAiMAX™, siPORT™ NeoFX™ and DharmaFECT 4. PTK9 downregulation was significant in cells transfected with Pre-miR-1 and all Lipofectamine™ RNAiMAX™ volumes, Low and Medium siPORT™ NeoFX™ volumes and Low DharmaFECT 4 ($P < 0.01$). Low and Medium Lipofectamine™ RNAiMAX™ and Low DharmaFECT 4 produced the highest knock-down. Cells transfected with pmNC#1 and all Lipofectamine™ RNAiMAX volumes or Low and Medium siPORT™ NeoFX™ volumes showed an significantly increased PTK9 expression compared to the non-transfected control. The Non-Transfected Controls showed lower PTK9 expression when compared to the pmNC#1 samples ($P < 0.05$). The PTK9 expression in the pmNC#1 samples was set at 100%. All experiments were performed in biological triplicate. Values indicate the mean of relative expression ± SD. A t-test was used to compare sample means.

3.4.3 — Forward Transfection Optimisation of Pre-miR™ miRNA Precursors

Forward transfection was optimised for NTera-2 and 2102Ep cells as reverse transfection resulted in an inadequate reduction of PTK9. As stated in Section 2.2.1.3, an advantage of forward transfection was the ability to diagnose the state and confluence of the cells on the plate prior to transfection. This was advantageous as changes cell morphology could occasionally be observe if the cells were growing at a density which was too low.

3.4.3.1 — Optimisation of Experimental Media

To determine the effects of short-term incubation of 2102Ep and NTera-2 cells in Opti-MEM® I, the first step to optimise forward transfection for these cell lines was to test cell viability after a 6-hour incubation of these cells in Opti-MEM® I medium. This incubation period was chosen as it would represent the length of time that 2102Ep cells would undergo transfection before the transfection medium would be replaced by hEC growth medium. The sensitivity of 2102Ep cells to a 6-hour incubation in Opti-MEM® I medium was tested as described in Section 3.3.5.1. The results expressed in percentage viability are displayed in Figure 3.7.

Briefly, the cells were pre-seeded on 96-well plates 24 hours prior to exposure to experimental conditions. One sample set was exposed to Opti-MEM® I medium for 6 hours, while the hEC medium was refreshed on the second sample set. After this 6 hour incubation window, the media were replaced by hEC medium for another 48 hour incubation. The results show no significant difference in cell viability between the two controls. 2102Ep cells are therefore insensitive to serum-free Opti-MEM® I exposure for the preferred transfection period of 6 hours.

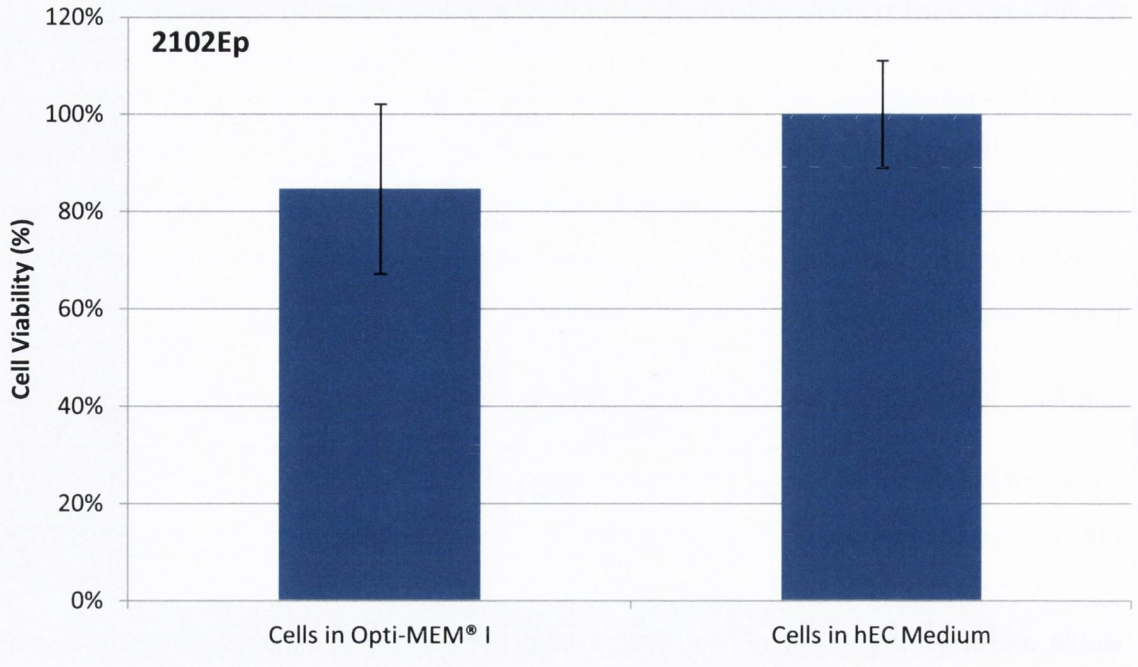


Figure 3.7—Cell viability of 2102Ep cells incubated in Opti-MEM® I for 6 hours versus cells exposed only to hEC medium. No significant difference in viability exists between cells grown for 72 hours in hEC medium and cells grown for 66 hours in hEC medium with a 6-hour Opti-MEM® I incubation period after the first 24 hours (P = 0.2430). All experiments were performed in biological triplicate. Cell viability of cells grown in hEC medium was set at 100%. Values indicate the mean cell viability ± SD. A t-test was used to compare sample means.

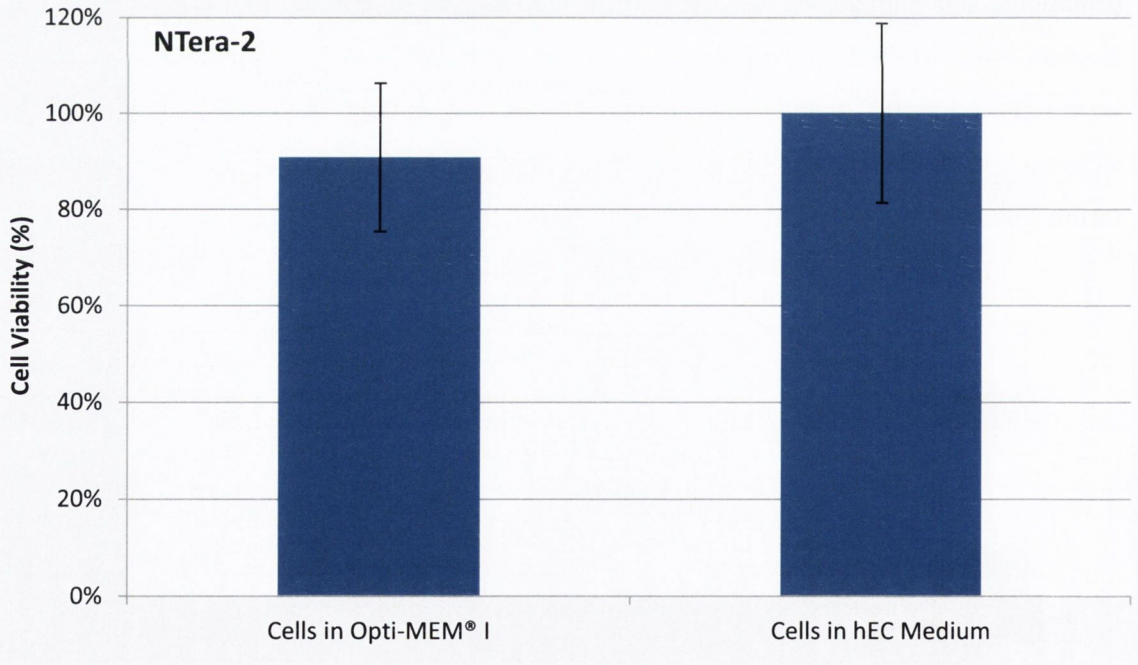


Figure 3.8—Cell viability of NTera-2 cells incubated in Opti-MEM® I for 6 hours versus cells exposed only to hEC medium. No significant difference in viability exists between cells grown for 72 hours in hEC medium and cells grown for 66 hours in hEC medium with a 6-hour Opti-MEM® I incubation period after the first 24 hours (P = 0.1330). All experiments were performed in biological triplicate. Cell viability of cells grown in hEC medium was set at 100%. Values indicate the mean cell viability ± SD. A t-test was used to compare sample means.

3.4.3.2 — Optimisation of Forward Transfections

For forward transfections in NTera-2 cells, only Lipofectamine™ RNAiMAX™ was included in the forward transfection optimisation stages. At the time of this experiment, Lipofectamine™ RNAiMAX™ was the only transfection agent to produce substantial PTK9 knock-down upon the reverse transfection of Pre-miR-1. However, based on improved methods, the results shown for reverse transfections were recalculated from the original raw data after transfection optimisations were completed. Previous calculations produced the results (not shown) which suggested Low and Medium Lipofectamine RNAiMAX to be the only viable choice for forward transfections. However, in hindsight, Low DharmaFECT 4 could have been included in this choice.

As reverse transfections with Lipofectamine™, siPORT™ NeoFX™ and DharmaFECT 4 were unsuccessful in 2102Ep cells, forward transfections with Lipofectamine™ 2000 were attempted. Lipofectamine™ 2000 was previously successfully used for forward transfections with this cell line.[288,333,334] The manufacturer recommends forward transfections with Lipofectamine™ 2000 as this transfection agent can be quite toxic to the cells. The official protocol for Lipofectamine™ 2000 recommends a cell confluency of 30-50% for siRNA transfections, but a higher confluency if high cytotoxicity is observed. Due to the naturally high mitotic rate of 2102Ep cells, a maximum confluency of 50% was warranted, as higher transfection confluency lead to complications due to an overconfluence of the culture. No reverse transfection experiments were attempted with Lipofectamine™ 2000.

As with reverse transfection optimisation, forward transfections in 2102Ep and NTera-2 cells were performed with 50 nM Pre-miR-1 and pmNC#1. The method used is described in Section 2.2.2.3. The results of PTK9 expression in 2102Ep and NTera-2 cells are displayed in Figure 3.9 and Figure 3.10 respectively.

Forward transfections of Pre-miR-1 with Lipofectamine™ 2000 yielded substantially better results in 2102Ep cells than did the reverse transfections with the three previous transfection agents. PTK9 expression was significantly reduced to 50% and 35% for cells transfected with Medium (0.25 µl/well) and High (0.375 µl/well) Lipofectamine™ 2000 concentrations respectively. High Lipofectamine™ 2000 was therefore a good candidate for functional transfections of Pre-miR™ miRNA Precursors.

Results show that highest PTK9 knock-down occurred in NTera-2 cells transfected with Pre-miR-1 using Medium (0.2 µl/well) and High Lipofectamine (0.3 µl/well) concentrations, with expression values of approximately 29-33%. This was considered sufficient for functional transfections of Pre-miR™ miRNA Precursors and below the maximum remaining PTK9 expression of 40%, considered

by Ambion to be indicative of high transfection efficiency.[328] Significant difference in PT9 expression existed in cells transfected with pmNC#1 and both non-transfected controls. This potentially could have been caused by transfection impact on PTK9 expression, however, as the fold difference between the pmNC#1 samples and non-transfected controls remained <2 and as this phenomenon was not demonstrated in the other results, this was not considered to be of great importance to further utilisation of the method.

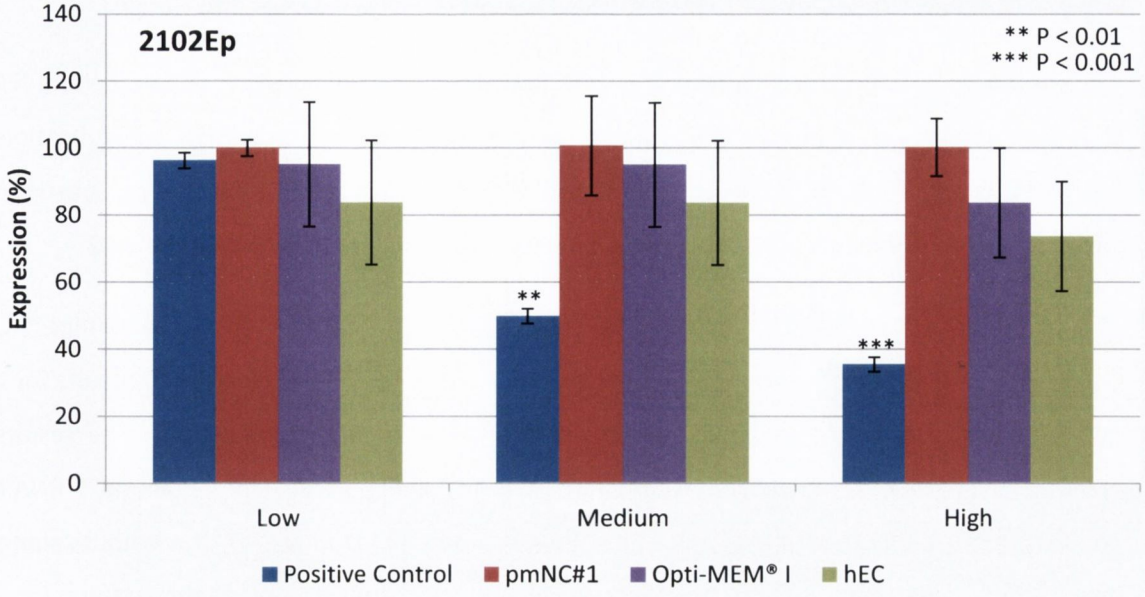


Figure 3.9—Percentage expression of PTK9 in 2102Ep cells forward transfected with miR-1 using three different volumes of Lipofectamine™ 2000. PTK9 expression in cells transfected with pmNC#1 (red) was set at 100%. PTK9 expression was significantly lower for 2102Ep cells transfected with 50 nM Pre-miR-1 and Medium ($P = 0.0042$) and High ($P = 0.0002$) Lipofectamine™ concentrations. No significant difference reported for cells transfected at Low Lipofectamine™ concentrations ($P = 0.1289$). No significant difference exists between pmNC#1-transfected cells and both non-transfected controls (all $P > 0.05$). The PTK9 expression in the pmNC#1 cells was set at 100%. All experiments were performed in biological triplicate and the values reflect the mean relative expression ± SD.

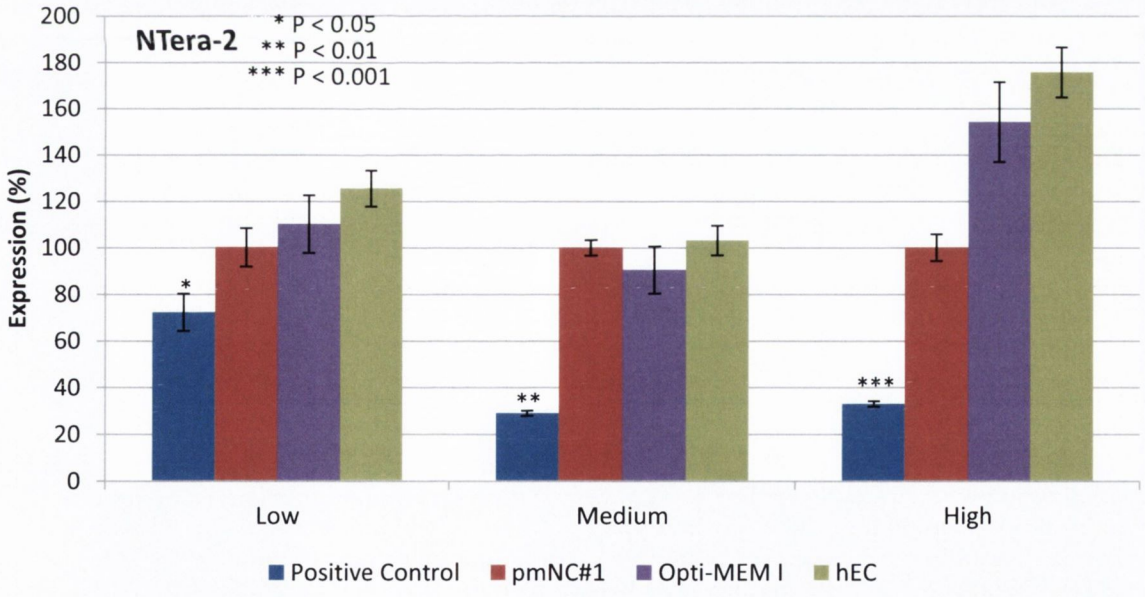


Figure 3.10—Percentage expression of PTK9 in NTera-2 cells forward transfected with miR-1 using three different volumes of Lipofectamine™ RNAiMAX. PTK9 expression was significantly lower in NTera-2 cells transfected with Pre-miR-1 at all Lipofectamine RNAiMAX concentrations compared to the pmNC#1 samples (Low: $P = 0.0138$, Medium and High: $P < 0.0001$). For cells transfected with pmNC#1 in High Lipofectamine™ RNAiMAX significant differences in PTK9 expression exist when compared to non-transfected controls in Opti-MEM® I and hEC ($P = 0.0068$ and $P = 0.0004$ respectively) The PTK9 expression in the pmNC#1 cells was set at 100%. All experiments were performed in biological triplicate and the values reflect the mean relative expression ± SD. A t-test was used to compare sample means.

3.4.3.3 — Cytotoxic Effects of Pre-miR™ Transfection

To assess impact of Pre-miR™ miRNA Precursor transfection on cell viability in 2102Ep and NTera-2 cells, transfection of pmNC#1 was performed at Low, Medium and High concentrations of Lipofectamine™ 2000 (2102Ep) or Lipofectamine™ RNAiMAX. Cell viability was subsequently assess as is described in Section 2.1.8. Results are shown in Figure 3.11 and Figure 3.12.

The transfection of Pre-miR™ miRNA Precursors in 2102Ep cells at all Lipofectamine™ 2000 volumes tested does not seem to affect the viability of the cells. The cell viability results for both cell lines are within a range which is sufficient for experimental transfections. The results for NTera-2 cells show no significant reduction in cell viability at all Lipofectamine™ RNAiMAX concentrations compared to the cells which were incubated in Opti-MEM® I for 6 hours alongside transfection. However, due to the large error for the latter samples, the results for cells transfected with Medium (0.2 µl/well) and High (0.3 µl/well) concentrations of Lipofectamine™ RNAiMAX were compared to those which were transfected with Low transfection agent concentrations (0.1µl/well). This showed a significant reduction of cell viability of 13% and 16% respectively. This reduction was considered acceptable for further experimentation as directed by the manufacturer and that which is commonly acceptable for transfections.

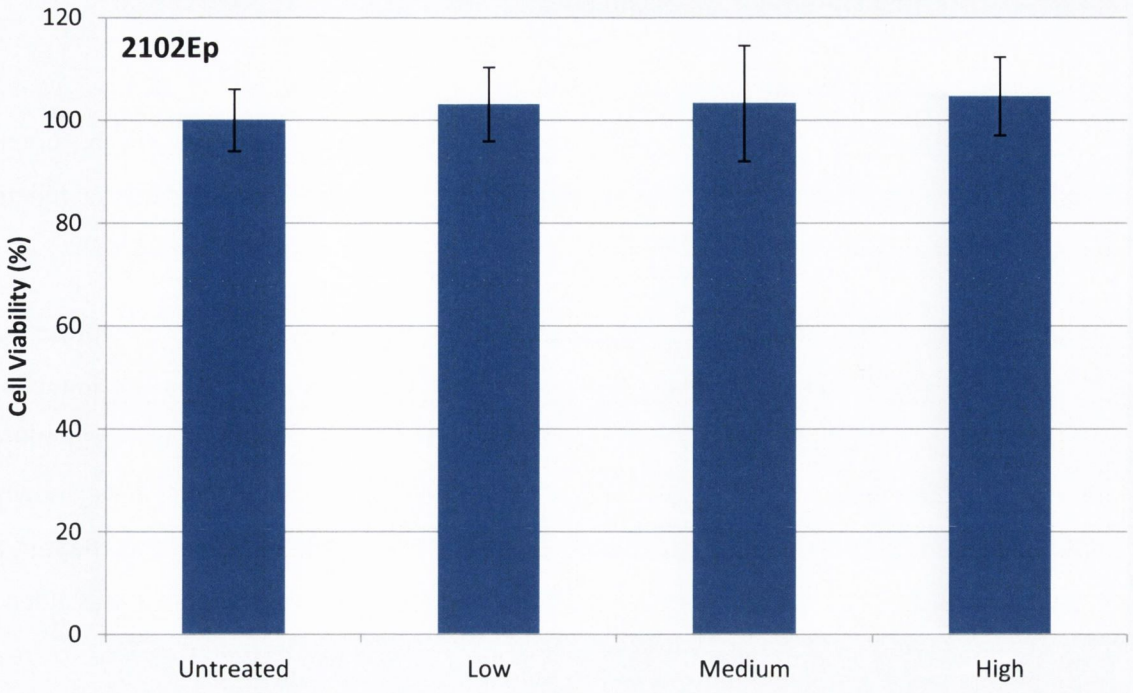


Figure 3.11—Cell viability of 2102Ep cells transfected with 50 nM pmNC#1 at different concentrations of Lipofectamine™ 2000. No significant difference of cell viability exists between the transfected cells and the untreated cells ($P > 0.05$). Cell viability of untreated cells was set at 100%. Values indicate the mean cell viability ± SD. A t-test was used to compare sample means.

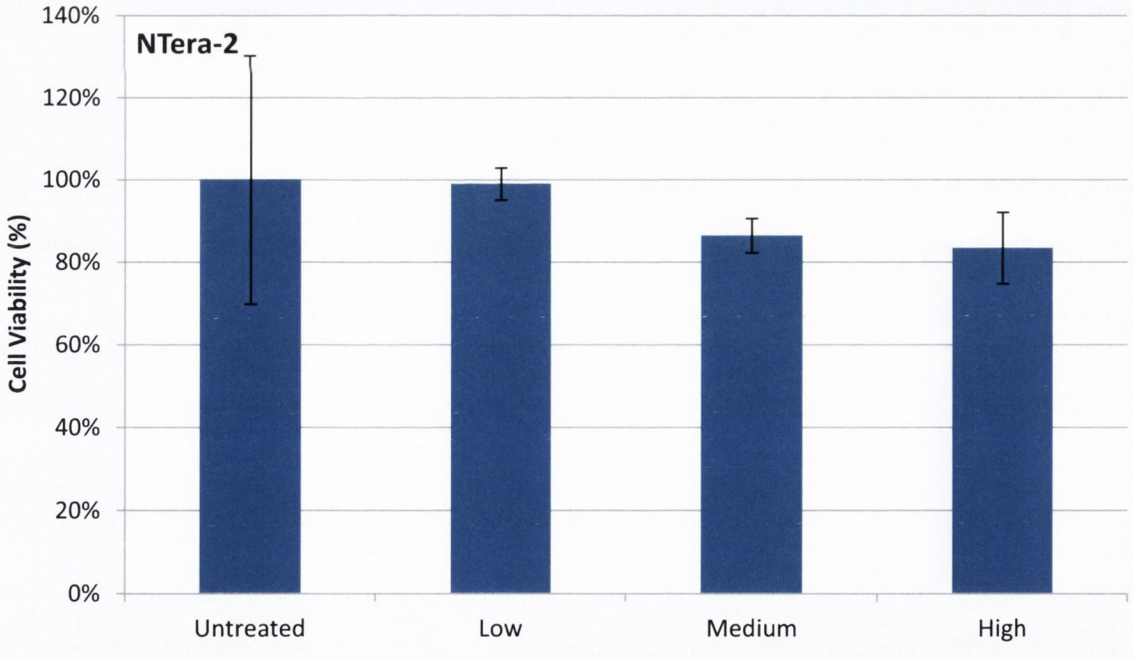


Figure 3.12—Cell viability of NTera-2 cells transfected with 50 nM pmNC#1 at different concentrations of Lipofectamine™ RNAiMAX. No significant difference exists ($P > 0.05$) between the transfected cells and cells incubated in Opti-MEM® I. Significant difference did exist between the cells transfected at higher Lipofectamine™ RNAiMAX concentrations and those transfected at Low (0.1 µl/well) Lipofectamine™ RNAiMAX concentrations (Medium vs Low: $P = 0.0147$, High vs Low: $P = 0.0305$). Cell viability of untreated cells was set at 100%. Values indicate the mean cell viability ± SD. A t-test was used to compare sample means.

3.4.4 — Functional Upregulation of miRNAs

Transfection optimisation of Pre-miR™ miRNA precursors in 2102Ep and NTera-2 cells was successfully completed and a laboratory protocol was established with which functional transfections could be conducted. The second part of this chapter consists of the functional upregulation of selected miRNA.

The overall aim of this thesis was to find mechanisms and pathways which are involved in CSC differentiation which could potentially be used to halt CSC proliferation by the induction of differentiation of these cells. One mechanism investigated in this study is the function of miRNA in the differentiation program of CSCs. As shown in Section 3.1, previous studies have shown the importance of miRNAs in the differentiation of ESCs and ECCs. In this Section of this chapter, a set of miRNA were chosen for overexpression in 2102Ep and NTera-2 cells. This choice was aided by a previously published study conducted by this laboratory.[274] The aim of this study was to create a list of miRNA which were deregulated in NTera-2 and 2102Ep cells upon their 3-day incubation in hEC medium supplemented with 10^{-2} mM all-trans retinoic acid. In total, 317 miRNAs were assayed in both cell lines, of which 113 were differentially regulated in NTera-2 cells and 59 were deregulated in 2102Ep cells. The set of five miRNAs chosen for overexpression studies were all significantly upregulated in differentiating NTera-2 cells and three were significantly upregulated in 2102Ep cells.

A total of five selected miRNA were functionally overexpressed in 2102Ep and NTera-2 cells in two separate stages. The reason for this division was the fundamental postulation on which these miRNA were chosen as candidates. The first stage of miRNA overexpression studies were conducted with miR-10a, miR-99a and miR-184. These miRNAs were thought to play a role in hECC differentiation as they are all upregulated in both NTera-2 and 2102Ep cells upon their 3-day incubation in 10^{-2} mM ATRA. These miRNA were also the only three in the data set to exhibit converse upregulation in both cell lines. It was therefore hypothesised that these miRNA could play important parts in processes connected to ATRA-induced differentiation pathways.

The second set of miRNAs selected to be overexpressed consisted of miR-219 and miR-335. These were chosen using slightly different criteria compared to miR-10a, miR-99a and miR-184. A list of miRNA upregulated >2-fold in NTera-2 cells after 3-day ATRA-induced differentiation (as generated by a previous study in the laboratory) was compiled.[274] The conserved targets of each miRNA on this list were predicted *in silico* with the TargetScan 5.1 miRNA target prediction software.[191,192,329] Cai et al. compiled a validated list of 55 genes claimed to be solely expressed in hESCs compared to differentiated derivatives.[335] This list was obtained by a massively parallel

meta-analysis of signature sequencing, expressed sequence tag scan and microarray data and subsequent genomic and proteomic validation conducted in the study. The predicted targets generated from TargetScan 5.1 were filtered against the gene list generated by Cai et al. This was performed with the objective of generating a pruned list of miRNAs which target hESC markers and could potentially have roles in the suppression of these markers, contributing to the differentiation cascade of hESCs. The top ten upregulated miRNA from the pruned list of miRNAs with their potential target hESCs markers is displayed in Table 3.8. The full list can be found in Supplemental Table 1. Further qualitative selection was performed by focusing on the highest upregulated miRNAs. MiR-10a has already been tested to no result in this study. Let-7c, a member of the let-7 family of miRNAs has already been extensively characterised in hESC, ECC and CSC differentiation as has its target, LIN28.[336–339] As the goal of this study was to search for novel miRNAs of interest to ECC differentiation, this miRNA was not included in further testing. MiR-219 and miR-335 were both among the highest upregulated miRNAs after a 3-day, ATRA-induced differentiation of NTera-2 cells. However, no expression increase was detected in 2102Ep cells. A further look at the potential targets of miR-335 and miR-219 further suggested a role in hESC and ECC differentiation, especially for miR-335 as it potentially conservatively targets crucial stem cell marker OCT4 (POU5F1). Further prediction analysis conducted with the widely used miRANDA miRNA target prediction software package suggested, besides OCT4, SOX2 as another candidate target of miR-335.[322,340,341] SOX2 was also targeted by miR-219 as predicted by the miRANDA software. Its other potential targets of interest, NUMB and PODXL are both expressed in NTera-2 cells.[239,318,342] Briefly, NUMB is well characterised in development and is considered a determinant of stem cell fate.[343] It also has suggested tumour suppressive roles linked to CSCs.[343] PODXL gained added interest as this hESC and hECC marker is overexpressed in malignant testicular tumours and controls NTera-2 cell invasion.[239] With the background information on miR-219 and miR-335 the it was decided to functionally overexpress these miRNA in 2102Ep and NTera-2 cells.

Table 3.8—List of miRNAs upregulated in NTera-2 cells upon 3-day exposure to 10^{-2} mM ATRA which potentially target crucial hESC markers.

miRNA	Fold ↑ in NTera-2	Predicted Targets
miR-10a	∞	*NR5A2, SMAD2*
let-7c	∞	*DIAPH2, LIN28, PMAIP1, SMAD2*
miR-335	26.0	*POU5F1*
miR-219-5p	21.9	*NUMB, PODXL*
miR-509-3p	18.3	*PTEN, SMAD2*
miR-188-5p	8.3	*LIN28, SMAD2*
mir-140-3p	6.1	*PTEN*
miR-33a/b	6.0	*LIN28*
miR-1	5.9	*GJA1, SMAD2*
miR-18b	4.7	*KIT, LIN28, SMAD2, IGF2BP2*

Table 3.9 describes the expression pattern and fold-deregulation in 2102Ep and NTera-2 cells during the incubation of these cell lines in 10^{-2} mM ATRA for three days. Further investigation in previously published research was conducted to elucidate the function of these miRNAs and their potential contribution to hECC differentiation.

Table 3.9—MiRNA expression in NTera-2 and 2102Ep cells exposed to 10^{-2} mM ATRA. (adapted from Gallagher *et al.*[274])

miRNA	NTera-2	2102Ep
hsa-miR-10a	Exclusively expressed in differentiating cells	↑ 119.2-fold
hsa-miR-99a	↑ 22.5-fold	↑ 2.0-fold
hsa-miR-184	↑ 3.7-fold	↑ 3.8-fold
hsa-miR-219	↑ 21.9-fold	No change
hsa-miR-335	↑ 26.0-fold	No change

The previously reported defined and validated functions of miR-10a, miR-99a, miR-184, miR-219 and miR-335, which are considered relevant to cancer, stemness and embryonic development are stated in Table 3.10.

Functional miRNA overexpression for RNA expression studies were performed by transfecting Pre-miR™ miRNA precursors on 6-well plates to enable easy observation during growth and to obtain sufficient RNA and protein after cell lysis. The protocol for Pre-miR™ miRNA Precursor transfection on 6-well plates can be found in Section 3.3.6. As, at the time, no validated targets could be used to assess transfection efficiency with certainty, the relative expression of the miRNAs themselves were measured. This method of measuring transfection efficiencies of Pre-miR™ miRNA Precursors was conducted in previous research and successfully tested for this study by transfecting Pre-miR-1 in NTera-2 cells and measuring miR-1 expression with qRT-PCR as described in Section 2.5.2.3 (data not shown).[212,344,345] In addition to measuring miRNA expression, the expression of the three gene markers of pluripotency, OCT4, SOX2 and NANOG, were measured. These markers, detailed in Section 1.2 are crucial to maintaining the stem cell and self-renewal state of ESCs and ECCs. Their reduction are indicative of the occurrence of spontaneous differentiation of these cells and thus they can be used to determine whether differentiation is commencing. To further assess the presence of differentiation, three differentiation markers were also included in this study. These were GATA6, an endodermal/trophectodermal marker, NCAM1, an ectodermal marker and ENO3, a mesodermal marker. These markers were used for the purpose of determining differentiation in a previous study performed by this laboratory.[274] These differentiation markers were only used for 2102Ep and NTera-2 cells transfected with Pre-miR™ precursors for miR-10a, miR-99a and miR-184.

Table 3.10—Previously discovered functions of miR-10a, miR-99a and miR-184, miR-219 and miR-335.

miR-10a	- Tumour suppressor in chronic myeloid leukaemia by potentially inhibiting the expression of USF2.[344] - Enhances ribosomal protein translation and may positively control global protein synthesis.[213] - Can induce apoptosis in a TRAIL-independent manner.[346] - A nuclear associated miRNA in colon cancer cells.[347] - Endodermal differentiation marker during hESC differentiation.[348] - Binds EGR1 in PMA-induced K562 cells.[349] - Inhibited by RA-receptor antagonist and promotes metastatic behaviour of pancreatic tumour cells.[350] - Overexpressed in acute myeloid leukaemia with NPM1 mutations and targets MDM4.[351] - Contributes to neuroblastoma cell differentiation by targeting NCOR2 and SFRS1.[352,353] - Contributes to ATRA-mediated smooth muscle differentiation.[354] - Targets HOXD4 in breast cancer cells.[190]
miR-99a	- Lower expression correlates with lower survival in hepatocellular carcinoma and its overexpression inhibits tumour growth.[355] - Downregulated in cisplatin resistant embryonal carcinoma cells.[356] - Targets mTOR and FGFR3 and is downregulation by c-Src activation.[357] - Downregulated in advanced prostate cancer cell lines, inhibits their growth and reduces PSA expression by targeting SMARCA5, SMARCD1 and mTOR.[358]
miR-184	- Is specifically localised in embryonic development.[359–361] - Suppressor of neuroblastoma by targeting AKT2.[362,363] - Regulates neural stem cell proliferation and differentiation by targeting MBD1.[364] - Suppresses growth of glioma cell lines by downregulating AKT2 among other potential candidate genes.[345]
miR-219	- Promotes oligodendrocyte differentiation by targeting SOX6, HES5, PDGFRα, FOXJ3 and ZFP238.[365,366]
miR-335	- Suppresses gastric cancer metastasis by targeting BCL2L2 and SP1[367] - Breast cancer suppressor by targeting ERα, IGF1R, SP1 and ID4, all involved in the BRCA1 pathway.[368] - Oncogenic miRNA in astrocytoma by targeting DAAM1 and its suppression results in reduced tumourgenicity.[369] - Activates Wnt signalling, promoting osteogenic differentiation by targeting DKK1.[370] - Silenced in breast cancer and characterised as a robust inhibitor of tumour reinitiation.[371] - Inhibitor of reparative mesenchymal stem cell phenotype by targeting RUNX2.[372] - Suppresses breast cancer metastasis by targeting SOX4 and TNC.[373]

3.4.5 — Induced Overexpression of Five miRNAs in 2102Ep cells

In this Section, the results from miRNA overexpression in 2102Ep are shown. As mentioned in Section 3.4.4, because of different approaches, the miR-10a, -99a and -184 functional overexpression experiment and the miR-219 and miR-335 overexpression experiment were performed separately. Although modest alterations exist between the experiments, the results all functional miRNA experiments are shown together. Briefly, miRNA were functionally upregulated using the protocol in Section 3.3.6. After transfection any changes in cell morphology were recorded. Molecular analysis was performed by measuring miRNA expression and pluripotency marker expression. Differentiation marker expression was measured (for cells transfected with Pre-miR-10a, -99a and -184 only). The same samples were used for all aforementioned experiments.

3.4.5.1– Phenotypic Analysis

During the post-transfection incubation period images were taken of the cells to observe any morphological change. When differentiation is occurring, certain typical morphological changes to the cells are expected. As a model of hEC differentiation, during ATRA-induced differentiation of NTera-2 cells, phenotypic changes to the cells normally become apparent after 2-3 days.[374] The cells tend to flatten and grow larger due to an increase in cytoplasm. Their edges also become less defined. In addition to changes in cell morphology, lower confluence with the absence of apoptopic bodies and dead cells could indicate senescence, an ensuing feature of differentiation. However, as 2102Ep and NTera-2 cells tend to display some of these features spontaneously when grown at low densities, it is sometimes difficult to distinguish subtle induced morphological change from these unintended phenotypes.

The images of 2102Ep cells with increased miR-10a, -99a and -184 expression are shown in Figure 3.13. On the second day of incubation after transfection, 2102Ep cells were approximately 90% confluent. The decision was made to split the cultures 1:2 to avoid over-confluence. The images in Figure 3.13 were taken on the third day of incubation. No phenotypic change can be detected in the Pre-miR-10a transfected cells compared to pmNC#1-transfected cells. This is different for Pre-miR-99a and Pre-miR-184 transfected 2102Ep cells, which display a degree of flattening and enlargement compared to the pmNC#1-transfected cells. Furthermore, the latter cells appear more densely packed in colonies, whereas the functionally overexpressed cells display a more homogenous appearance. This is characteristic of differentiating ECC cells, however this cannot be confirmed until further research is conducted.

In the Figure 3.14 2102Ep cells are shown at 72 hours after transfection of Pre-miR-219 and Pre-miR-335. Little morphological differences between the Pre-miR-219 and Pre-miR-335 transfected 2102Ep cells and the pmNC#1-transfected cells could be detected.

In all cell culture plates, a slightly higher confluency was found in the vehicle and non-transfection control wells (not shown) indicating a slightly negative effect on cell viability due to the transfection of pmNC#1, however this was judged to be an unavoidable side-effect due to the requirement of high transfection agent concentration and medium cell confluence in order to maintain high transfection efficiency.

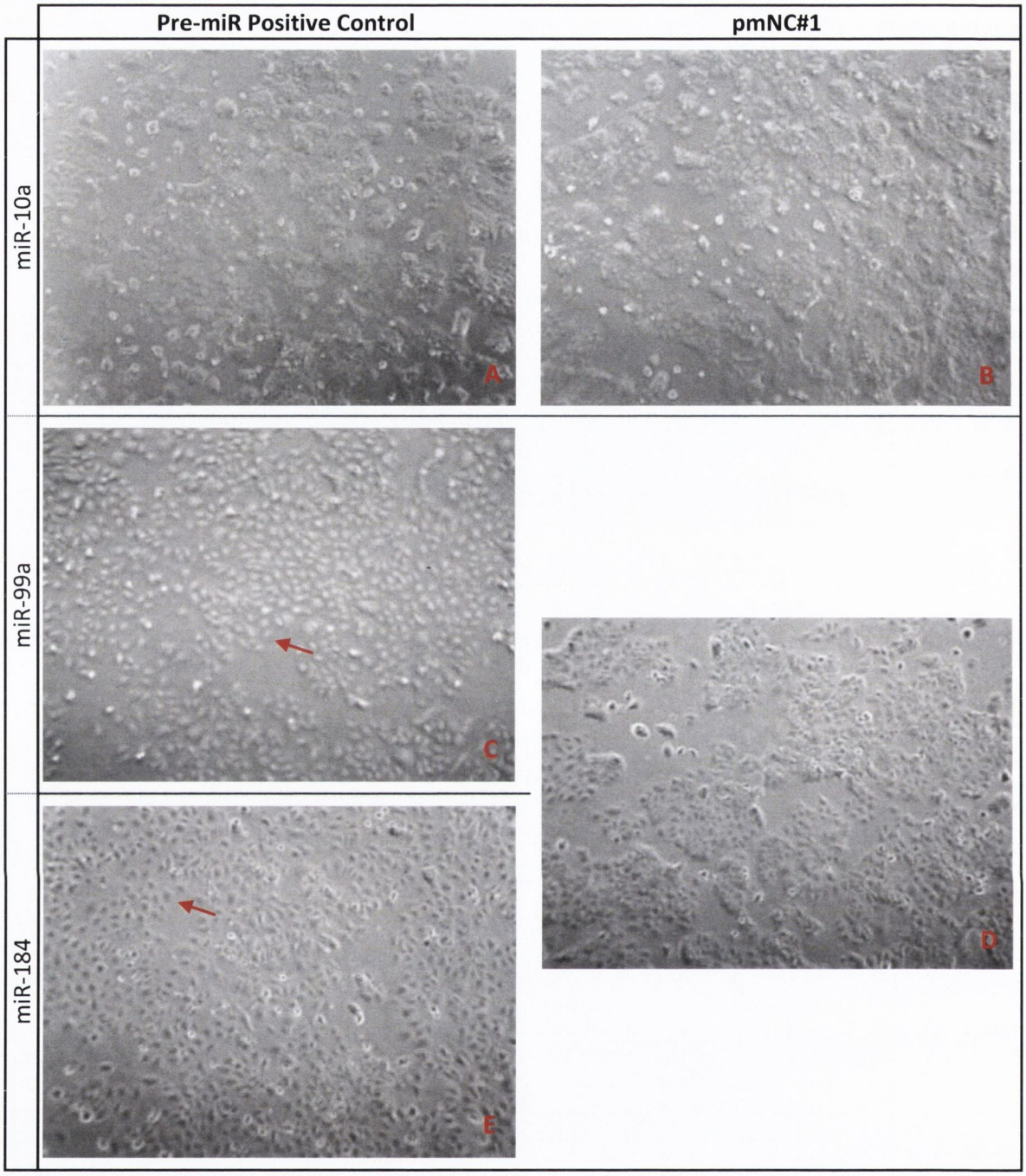


Figure 3.13—Morphology of 2102Ep cells transfected with Pre-miR-10a, -99a and -184 and incubated for three days. A more flattened, enlarged and less dense morphology was detected in samples transfected with Pre-miR-99a (C) and miR-184 (E), when compared to cells transfected with the pmNC#1 (D). Cells transfected with Pre-miR-10a (A) were unaltered. The arrows point towards examples. All images were photographed at a magnification of 50X.

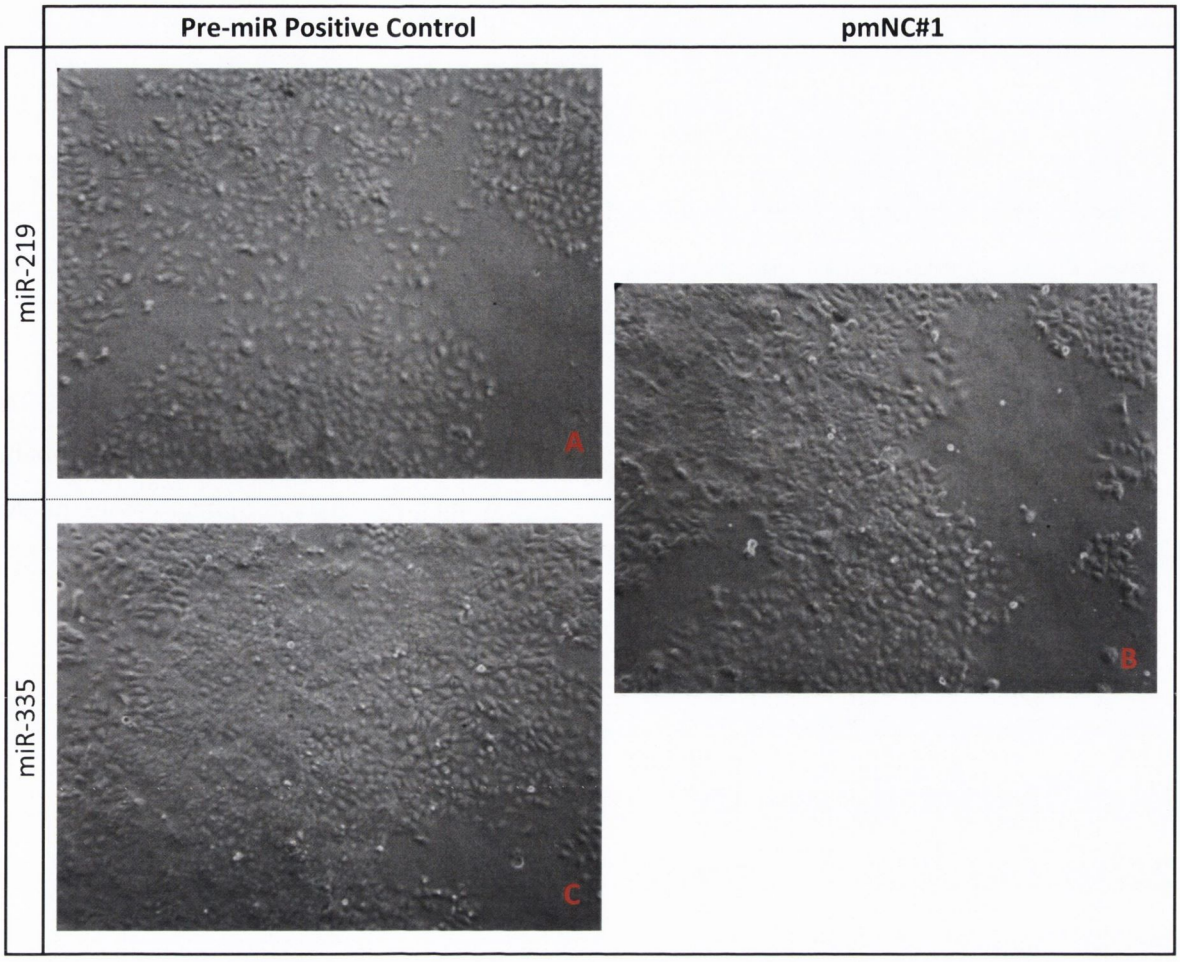
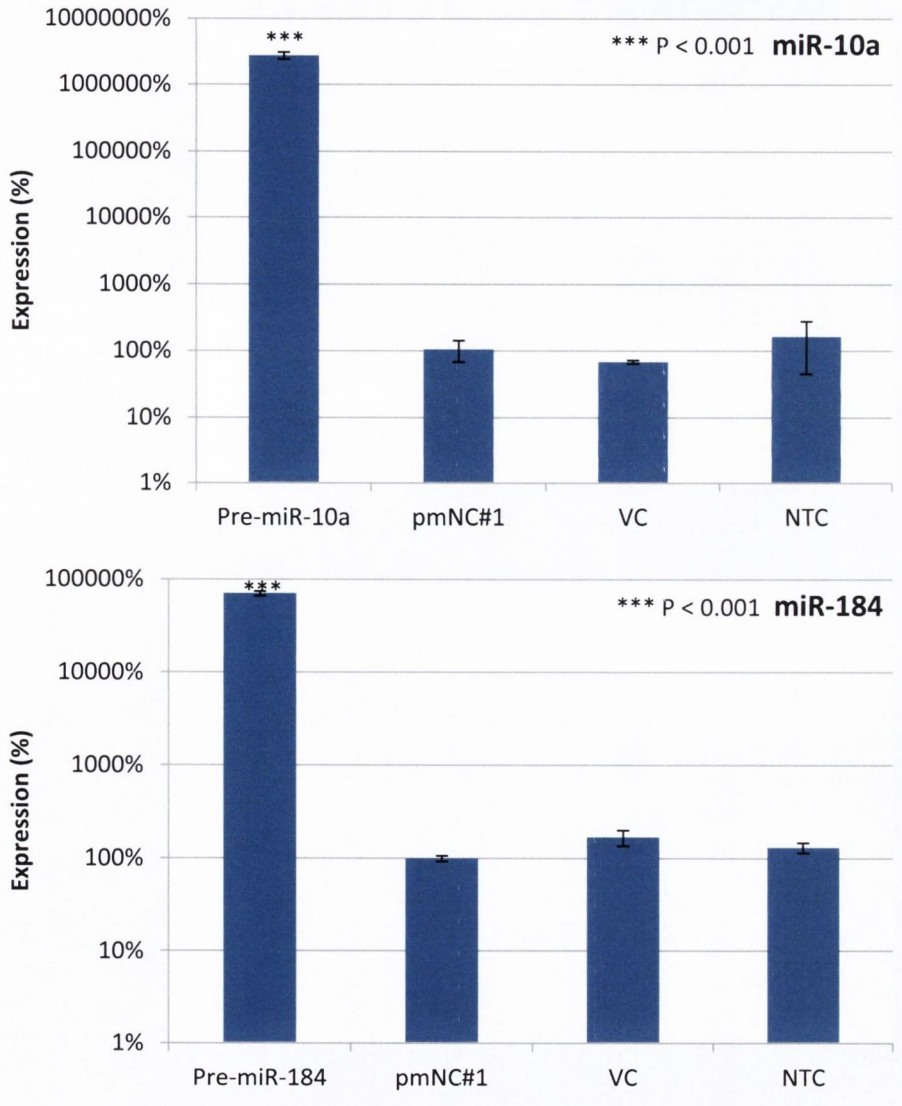


Figure 3.14—Morphology of 2102Ep cells transfected with Pre-miR-219 and Pre-miR-335 and incubated for four days. No notable morphological changes could be observed between the samples transfected with Pre-miR-219 (A) and Pre-miR-335 (C), compared to cells transfected with pmNC#1 (B). All images were photographed at a magnification of 50X.

3.4.5.2 — Expression of Functionally Upregulation miRNA

A strong upregulation of overexpressed miRNAs was expected in the samples transfected with the respective functional Pre-miR™ miRNA Precursors. MiRNA expression analysis was performed with the method described in Section 3.3.7. The miRNA expression results for Pre-miR-10a, -99a and -184 transfected 2102Ep cells are displayed in Figure 3.15. The results for Pre-miR-219 and Pre-miR-335 transfected cells are shown in Figure 3.16.

In all functionally transfected samples, miRNA upregulation is very substantial and significant. The target miRNAs are overexpressed by a factor of approximately 250-30,000, although this probably depends largely on original miRNA expression. Overall, the results showed that successful miRNA overexpression was achieved.



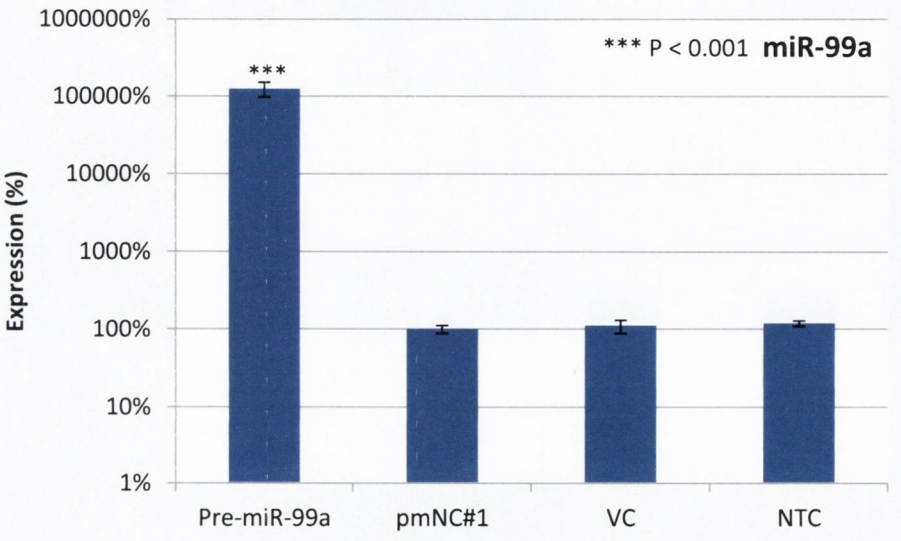


Figure 3.15—MiR-10a, miR-99a and miR-184 expression in functionally transfected 2102Ep cells. 2102Ep cells were transfected with Pre-miR™ miRNA precursors and incubated for 3 days. MiRNA expression, normalised to miR-16 expression is plotted on a logarithmic scale to account for high expression differences. In all experimental samples, specific miRNA expression was dramatically and significantly increased compared to the pmNC#1-transfected cells ($P < 0.0001$). MiR-10a and miR-99a expression in this negative control does not significantly differ from expression in the vehicle (VC) and non-transfected controls (NTC). However miR-184 expression in the pmNC#1-transfected cells significantly differs from that in the vehicle control ($P = 0.0236$) or the non-transfected control ($P = 0.0382$), however the vehicle and negative control, when compared to each other, do not ($P = 0.1448$). These differences are very small. The expression in the pmNC#1 samples was set at 100%. All experiments were performed in biological triplicate and the values reflect the mean relative expression ± SD. To compare sample means, a t-test was used.

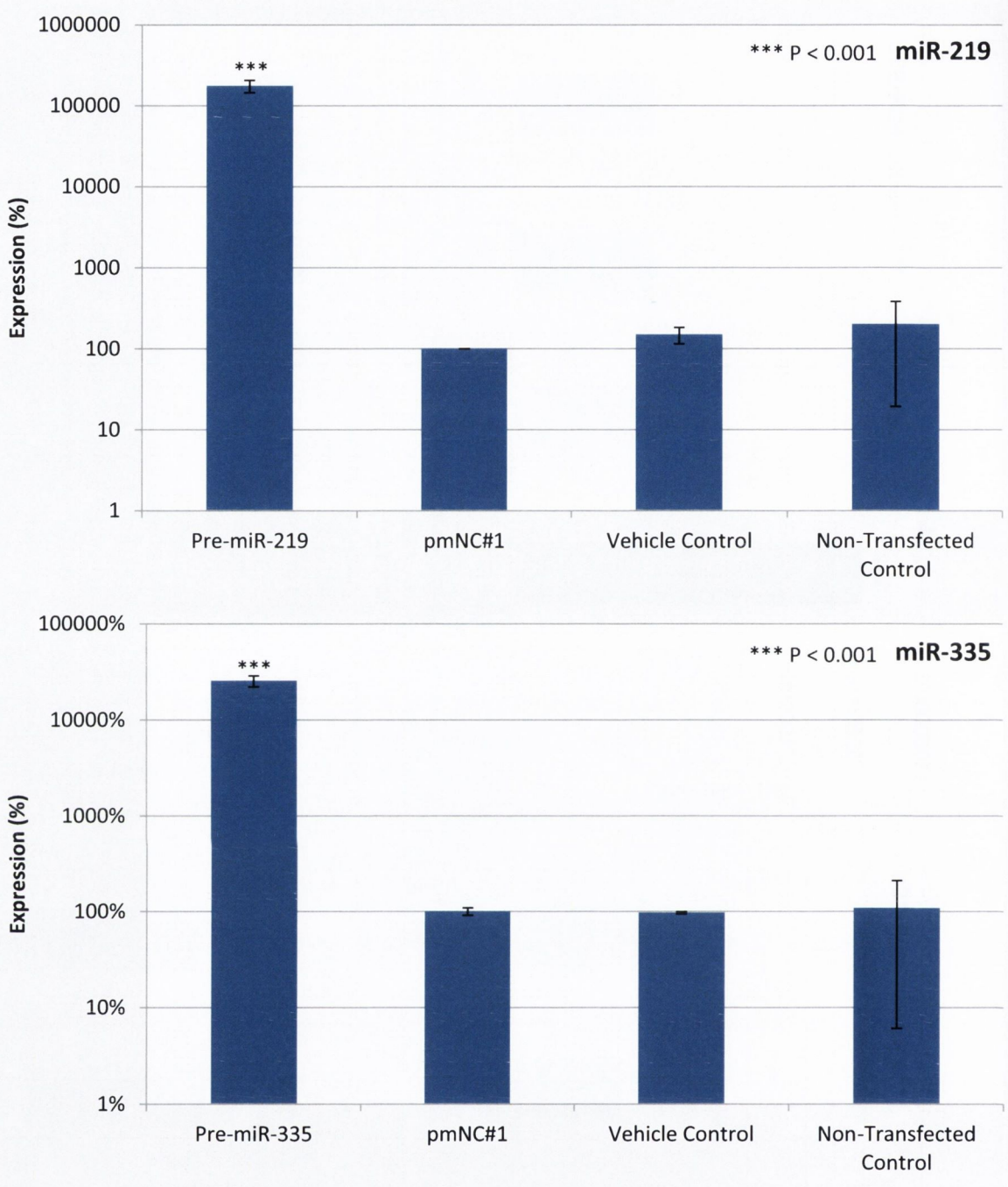


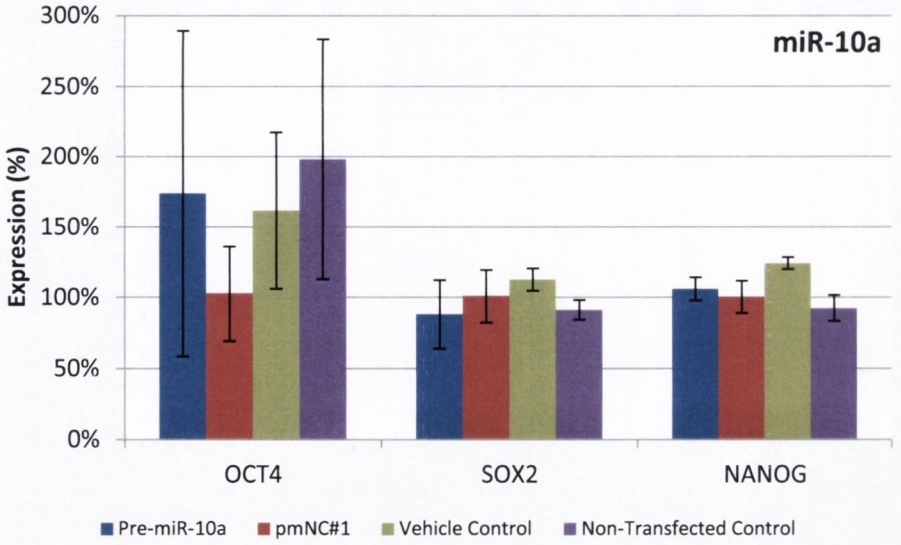
Figure 3.16—MiR-219 and miR-335 expression in functionally transfected 2102Ep cells. 2102Ep cells were transfected with Pre-miR™ miRNA precursor and incubated for three days. MiRNA expression, normalised to miR-16 expression is plotted on a logarithmic scale to account for high expression differences. In both experimental samples, specific miRNA expression was dramatically and significantly increased compared to the pmNC#1-transfected cells (miR-219: P = 0.0005 and miR-335: P = 0.0002). No significant differences exist among the pmNC#1 samples, vehicle control and non-transfected control. The expression in the pmNC#1 samples was set at 100%. All experiments were performed in biological triplicate and the values reflect the mean relative expression ± SD. To compare sample means, a t-test was used.

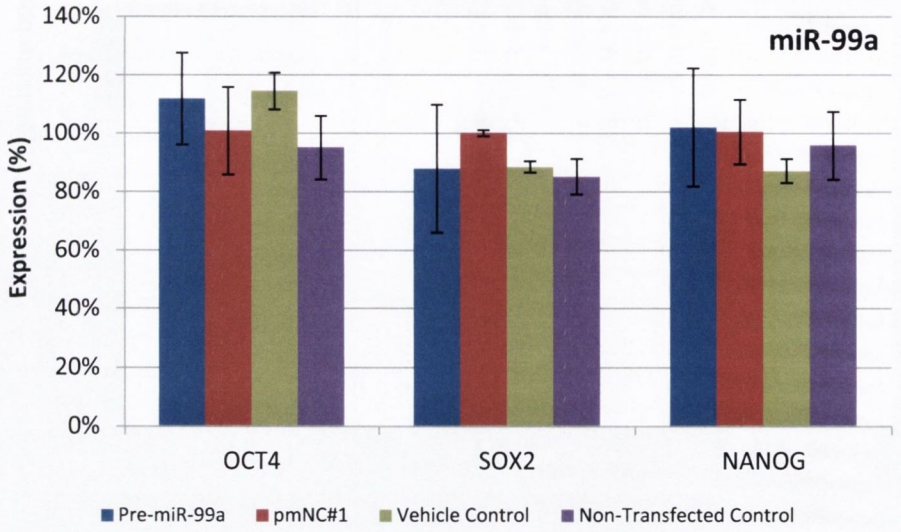
3.4.5.3 — Expression Analysis of Pluripotency Markers

As described in Section 1.2, the key pluripotency markers, OCT4, SOX2 and NANOG are crucial for the maintenance of self-renewal and the stem cell state in stem cells and ECCs. Loss of expression of these genes are commonly associated with the loss of pluripotency and with stem cell and ECC differentiation. SOX2, however, has been previously reported to maintain expression during early ATRA-induced NTera-2 differentiation due to its required functions in CNS cells (the cell fate of NTera-2 cells when induced by ATRA).[375,376] This phenomenon is further discussed in Section 3.5.

After transfection of specific Pre-miR™ precursors, the expression of key pluripotency markers in 2102Ep cells was measured with the method in Section 3.3.7. The results were displayed as a percentage expression relative to the pmNC#1-transfected cells. The TaqMan® assays used for measuring stemness marker expression are listed in Table 3.6. Pluripotency marker expression in 2102Ep can be found in Figure 3.14.

Upon inspection of the results, pluripotency marker expression in 2102Ep cells transfected with Pre-miR-10a, -99a, -219 and -335 showed no significant or substantial difference with the pmNC#1 samples. However, cells transfected with Pre-miR-184 showed a markedly and significantly lower expression of SOX2 and NANOG ($P < 0.0001$ and $P = 0.0175$ respectively), while OCT4 expression remained unchanged.





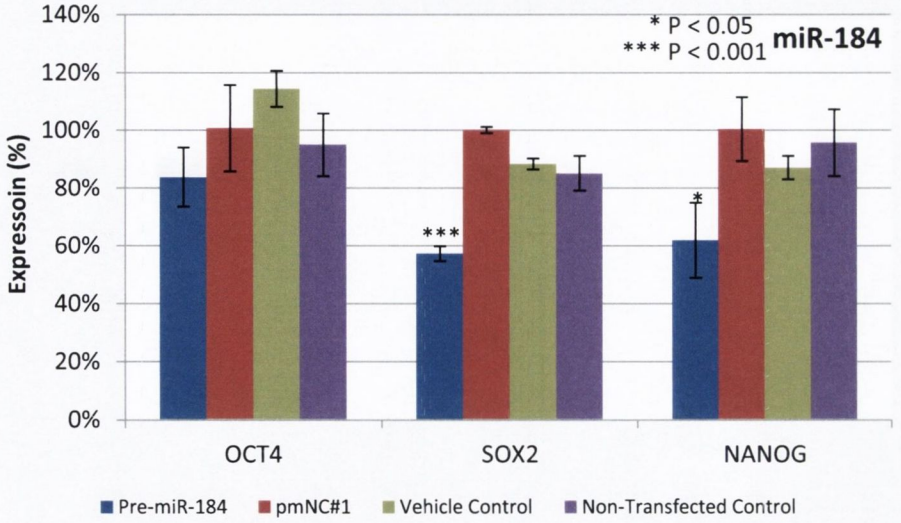


Figure 3.17—Pluripotency marker expression in 2102Ep cells transfected with Pre-miR-10a, -99a and 184. 2102Ep cells were transfected with 50 nM Pre-miR™ miRNA precursors and incubated for three days. Percentage gene expression is shown against the expression in the pmNC#1 samples, which was set at 100%. No significant alteration in pluripotency marker expression measured between cells transfected with Pre-miR-10a and Pre-miR-99a and their respective negative controls reported. SOX2 and NANOG expression in Pre-miR-184 transfected cells is significantly lower than the negative controls (SOX2: $P < 0.0001$, NANOG: $P = 0.0175$). OCT4 expression in Pre-miR-184 transfected cells did not significantly differ from the negative controls. All experiments were performed in biological triplicate and the values shown is mean relative expression of these triplicates ± SD. To compare sample means, a t-test was used.

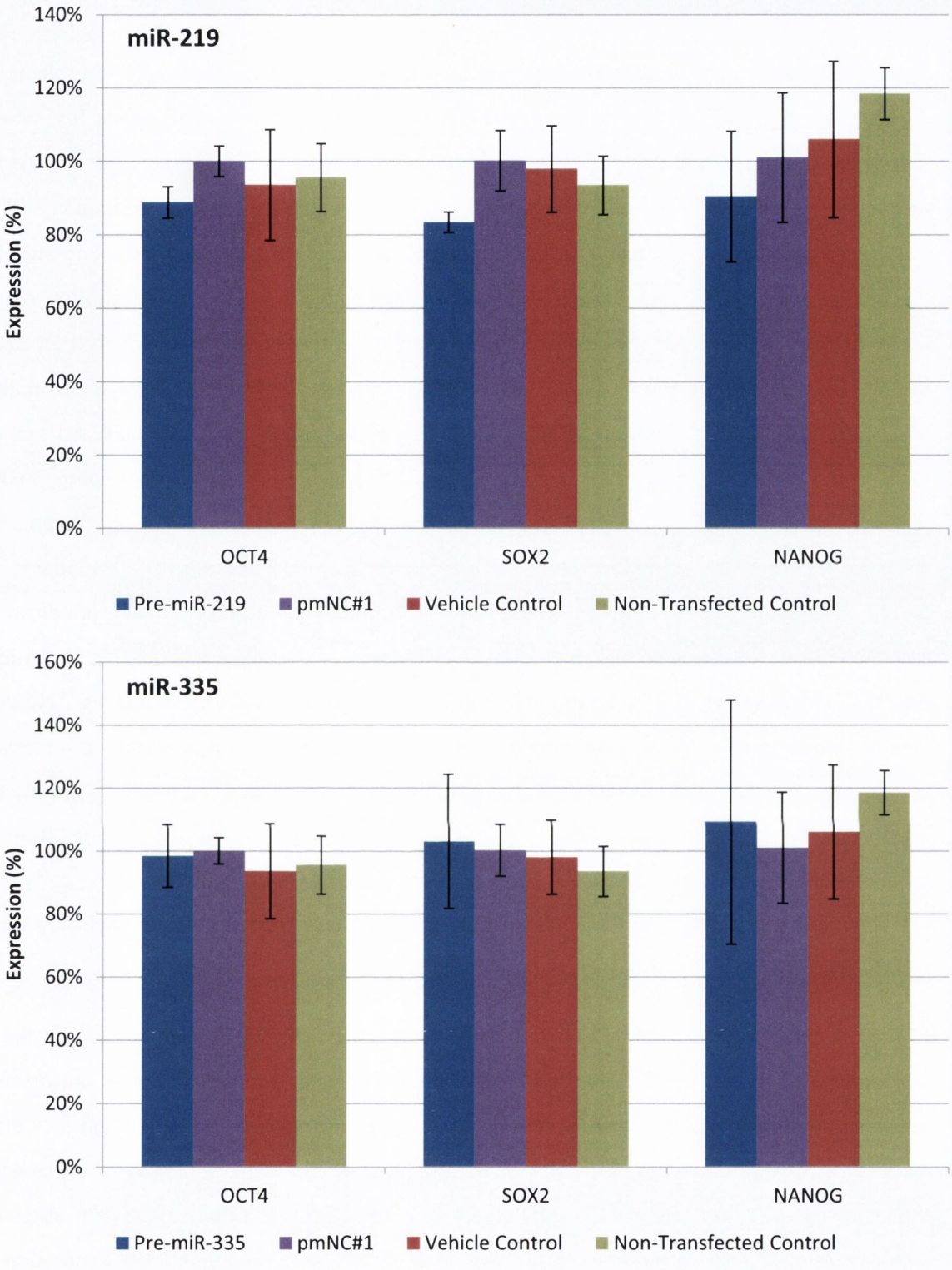


Figure 3.18—Pluripotency marker expression in 2102Ep cells transfected with Pre-miR-219 and Pre-miR-335. 2102Ep cells were transfected with 50 nM Pre-miR™ miRNA precursors and incubated for three days. Percentage gene expression is shown against the expression in the pmNC#1 samples, which was set at 100%. In cells transfected with Pre-miR-219, OCT4 and SOX2 expression was significantly, but unsubstantially, lower compared to the pmNC#1-transfected cells ($P = 0.0307$ and $P = 0.0287$ respectively), but this was not a substantial difference (<18%) and also was not significant when compared to the two other negative controls. No significant changes in pluripotency marker expression were measured in Pre-miR-335-transfected cells. All experiments were performed in biological triplicate and the values shown is mean relative expression of these triplicates ± SD. To compare sample means, a t-test was used.

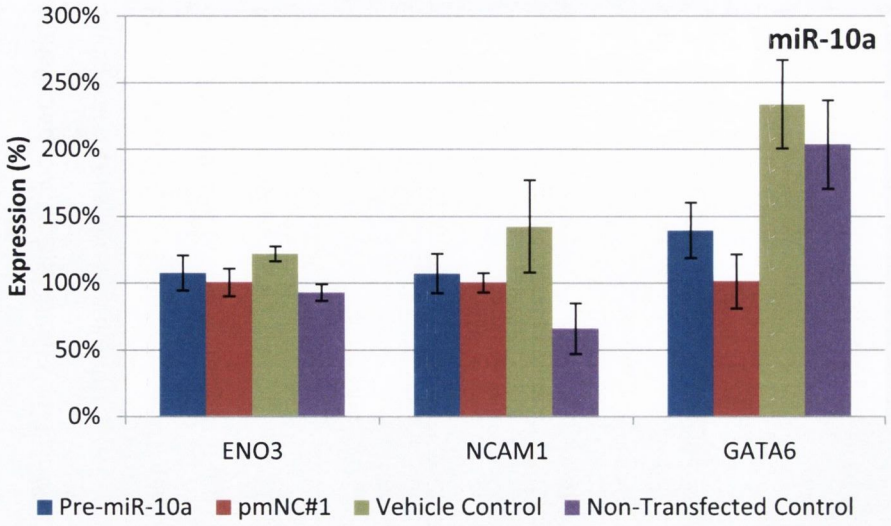
3.4.5.4 — Expression Analysis of Differentiation Markers

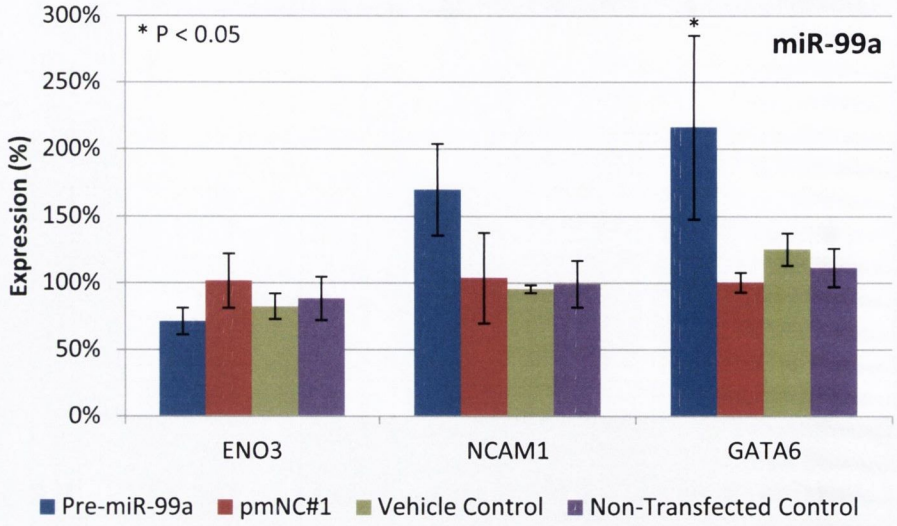
In 2102Ep cells transfected with Pre-miR-10a, -99a and -184, relative expression of a selected set of differentiation markers was measured. The relevance of this assessment was to test whether the expression of pluripotency markers, as shown in the previous Section, inversely matched the expression of the differentiation markers, thus confirming a degree of differentiation. In addition, this experiment would also test whether differentiation markers expressed independently from the pluripotency markers or if their expression is inversely linked. The differentiation markers were selected on the basis of their use in previous differentiation studies conducted on 2102Ep and NTera-2 cell lines in the laboratory. As NTera-2 cells are pluripotent, i.e. have the ability to differentiate in all three dermal layers, the three differentiation markers measured are each expressed in one of the dermal layers. The markers used were GATA6 (endoderm/trophectoderm), NCAM1 (ectoderm) and ENO3 (mesoderm). GATA6 has been shown to be expressed in the mesoderm, but is more commonly associated with the endoderm.[335,377] GATA6 has been widely used as a ESC and ECC marker of the primitive endoderm and trophectoderm as it has a crucial role in the development of these layers.[290,378–382] The promoter region of this gene can also be bound by the OCT4 and NANOG transcription factors.[79] NCAM1 is expressed during ectodermal and neurectodermal development and is widely used as a marker of ectodermal differentiation of adult stem cells, ESCs, IPSCs and ECCs.[383–388] ENO3 has rarely been used as a marker of differentiation, but has been shown to express during stem cell differentiation towards a muscle cell fate.[389] All three genes were upregulated during early ATRA-induced differentiation of NTera-2 cells, however ENO3 was reported to do this only by this laboratory.[274,318,390]

To measure the expression of differentiation markers, the batch of RNA from Section 3.4.5.2 was used. RNA was diluted to the required amount for all samples and qRT-PCR was performed as described in Section 2.5.2.1. Raw expression values were normalised to GAPDH and the relative expression was calculated using the $2^{-\Delta\Delta Ct}$ method. The results were displayed as a percentage expression relative to the pmNC#1-transfected cells. The TaqMan® Gene Expression assays used for measuring differentiation marker expression are listed in Table 3.6. Marker expression was normalised to GAPDH expression. Differentiation marker expression in 2102Ep cells is shown in Figure 3.19.

Differentiation marker expression significantly differs in 2102Ep cells transfected with Pre-miR-99a and Pre-miR-184. In Pre-miR-99a transfected cells, GATA6 expression showed a 2.2-fold increase compared to the pmNC#1-transfected cells. NCAM1 expression was 1.7-fold higher than the pmNC#1 samples. This is marginally insignificant when compared with expression in the

pmNC#1-transfected cells, but this is probably due to error as both the vehicle and non-transfected controls, which significantly differ, have similar NCAM1 expression. Pre-miR-184 transfected cells showed a substantial and significant 4.5-fold upregulation of GATA6. The miR-99a results reflect to some degree the phenotypic change observed in the functionally transfected cells. The miR-184 results reflect both the phenotypic change observed as the change in SOX2 and NANOG expression.





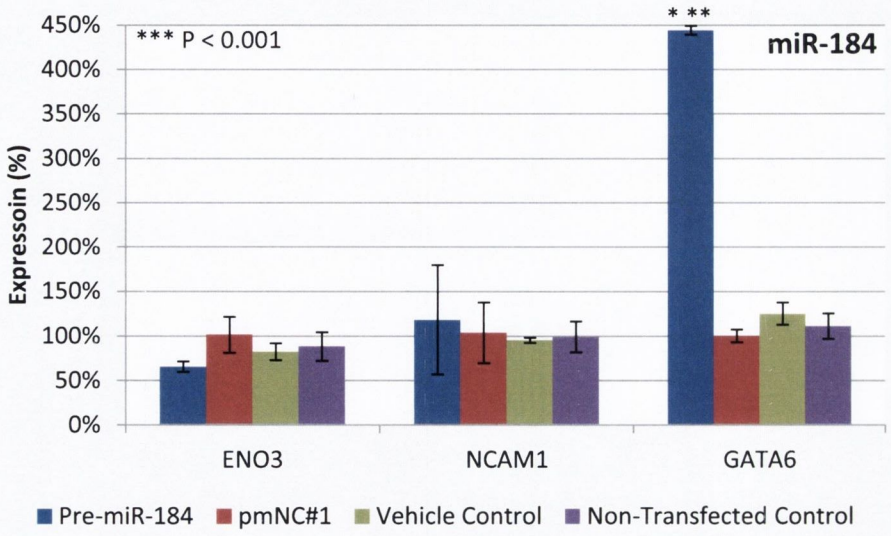


Figure 3.19—Differentiation marker expression in 2102Ep cells transfected with Pre-miR-10a, -99a and 184. 2102Ep cells were transfected with 50 nM Pre-miR™ miRNA precursors and incubated for three days. Percentage gene expression is shown against the expression in the pmNC#1 samples, which was set at 100%. No significant difference in differentiation marker expression between Pre-miR-10a transfected cells and pmNC#1-transfected cells could be determined. NCAM1 expression was significantly reduced in the non-transfected controls ($P = 0.0402$) and GATA6 expression was significantly increased in the vehicle and non-transfected controls ($P = 0.0140$ and $P = 0.0463$ respectively). In Pre-miR-99a transfected cells a significant increase of GATA6 expression was measured ($P = 0.0435$). A concurrent upregulation of NCAM1 was observed, however this was marginally insignificant when compared to the pmNC#1-transfected cells ($P = 0.0761$). When compare to the other two negative controls, a significant difference exists (vehicle control: $P = 0.0195$, non-transfected control: $P = 0.0328$). Significant upregulation of GATA6 was observed in Pre-miR-184 transfected cells ($P < 0.0001$), however, this did not concord with an increase in ENO3 and NCAM1 . ENO3 was, furthermore, significantly downregulated ($P = 0.0418$). All experiments were performed in biological triplicate and the values shown is mean relative expression of these triplicates ± SD. To compare sample means, a t-test was used.

3.4.6 — Induced Overexpression of Five miRNAs in NTera-2 Cells

In similar manner to functional Pre-miR™ miRNA precursor transfection, as described in Section 3.4.5, Pre-miR™ precursors were transfected into NTera-2 cells. As with 2102Ep cells, NTera-2 cells were transfected Pre-miR-10a, -99a and -184 in a single experiment and in a separate experiment NTera-2 cells were transfected with Pre-miR-219 and Pre-miR-335. These miRNA were selected for the reasons described in Section 3.4.4. After transfection, changes in cell morphology was analysed and molecular analysis was performed to assess miRNA expression and pluripotency marker expression. As with 2102Ep cells, differentiation marker expression was only measured in the NTera-2 cells transfected with Pre-miR-10a, -99a and -184. The same samples were used for all aforementioned experiments.

3.4.6.1- Phenotypic Analysis

Morphology of the NTera-2 cells transfected with Pre-miR-10a, -99a and -184 was examine under a light microscope at 50X and images were taken at three days after transfection as shown in Figure 3.20. NTera-2 cells transfected with Pre-miR-10a, Pre-miR-99a and Pre-miR-184 all showed no apparent morphological change compared to the pmNC#1 samples.

The images shown in Figure 3.21 were taken four days after Pre-miR™ miRNA Precursor transfection after the cells were split 1:2 at 72 hours. NTera-2 cells had a strong response to miR-335 upregulation, but no morphological change in Pre-miR-219 transfected cells were obvious. Cell confluence was much lower in the Pre-miR-335 transfected cultures compared to the pmNC#1 samples. In addition, the cells appeared flatter and larger. Some cells had an altered shape and displayed small pseudopodic projections. These features are characteristic of differentiation. No apoptotic bodies were found in the cultures and cells were counted to measure the ratio of dead and alive cells in each control (see Section 2.1.7 for protocol). No significant differences in ratios were found, suggesting a possible occurrence of cell senescence. No morphological deviation could be observed in NTera-2 cells transfected with Pre-miR-219.

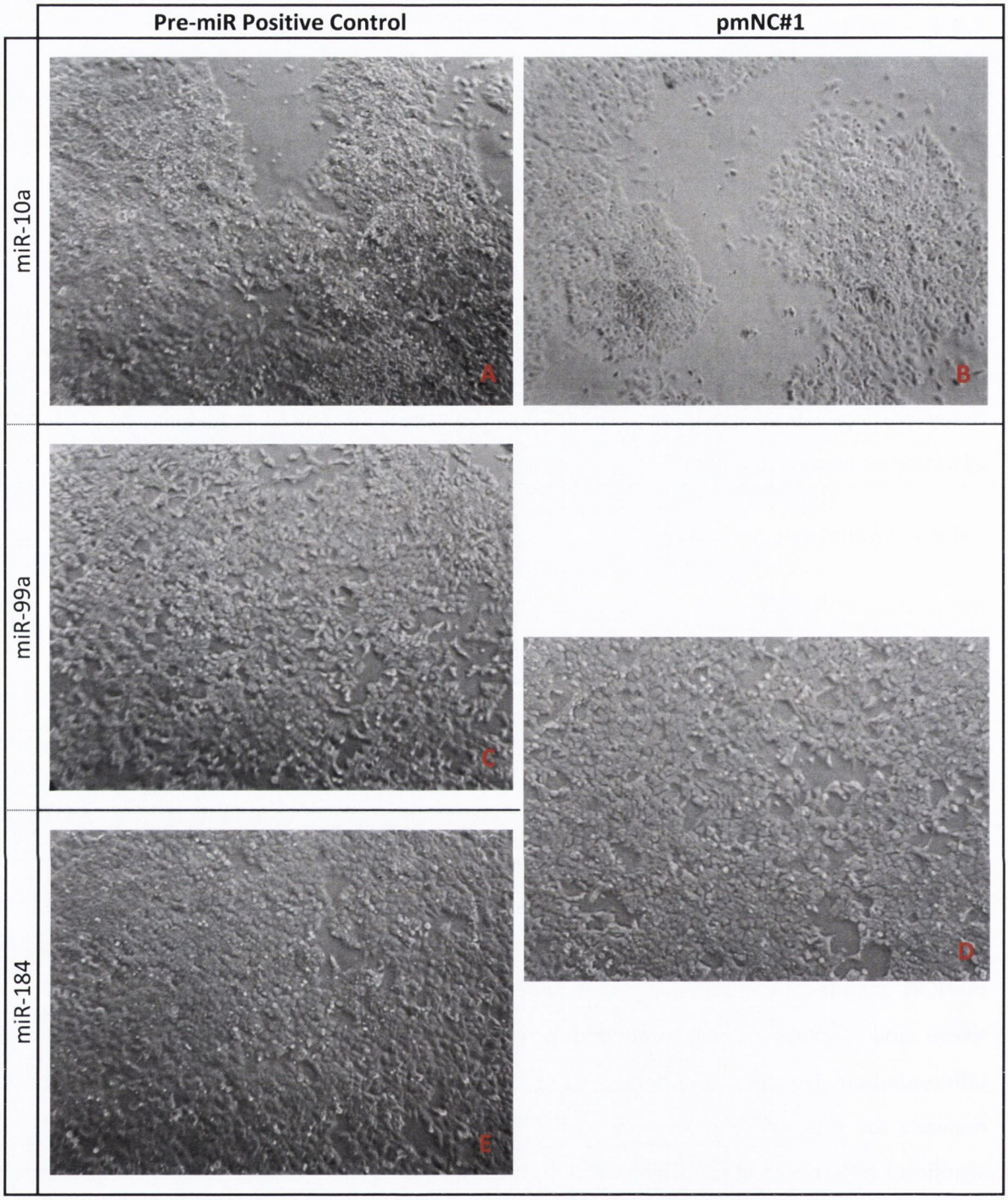


Figure 3.20—Morphology of NTera-2 cells transfected with Pre-miR™ precursors and incubated for three days. No notable morphological changes could be observed between the samples transfected with Pre-miR-10 (A), Pre-miR-99a (C) and Pre-miR-184 (E), compared to cells transfected with pmNC#1 (B and D). All images were photographed at a magnification of 50X.

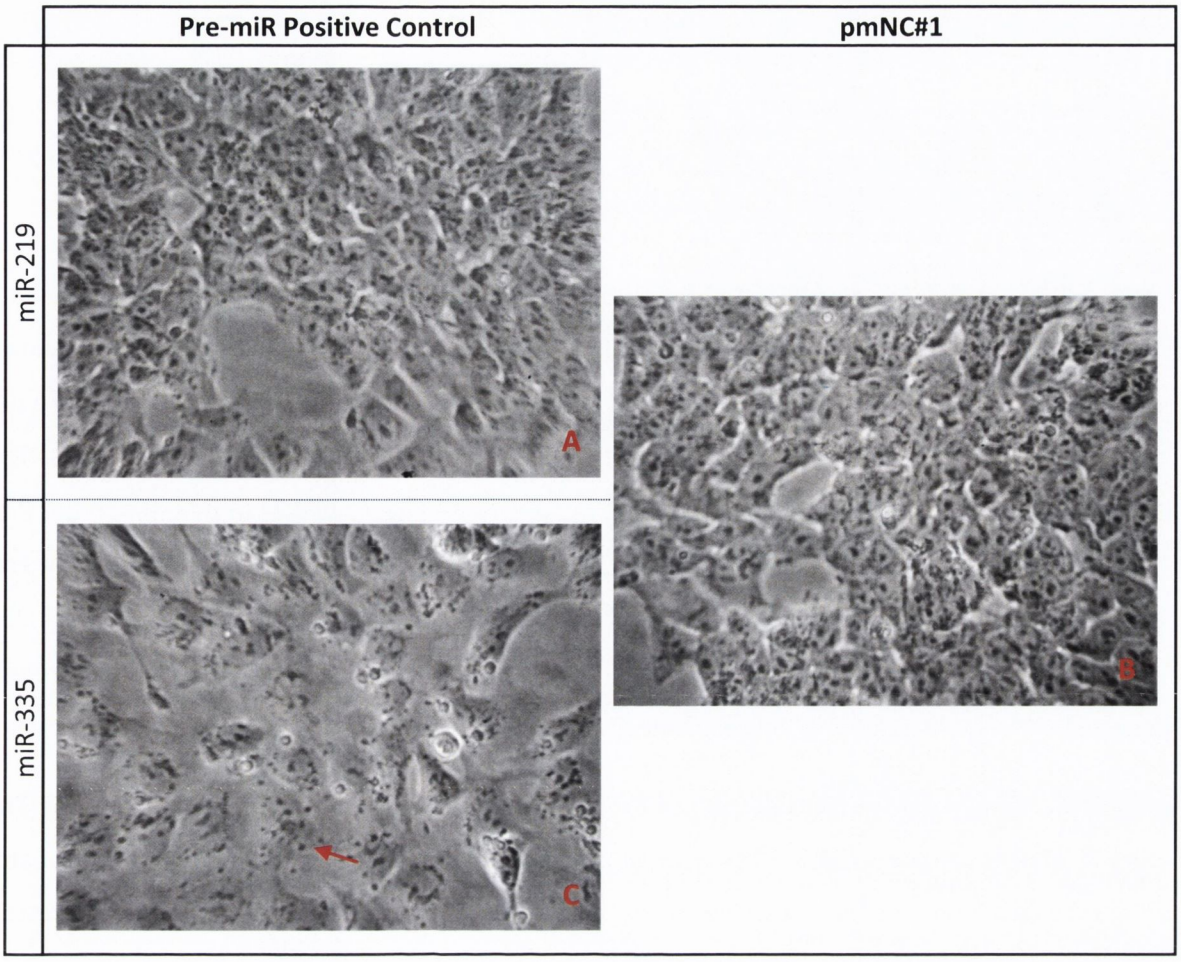


Figure 3.21—Morphology of NTera-2 cells transfected with Pre-miR™ precursors and incubated for four days. Pre-miR-335 transfection caused the cells to dramatically change towards a flattened, enlarged and less dense morphology (as indicated by the red arrow, C) compared to those transfected with pmNC#1 (B). The morphology of samples transfected with Pre-miR-219 did not change (A). All images were photographed at a magnification of 320X.

3.4.6.2- Expression Analysis of Functionally Upregulated miRNA

MiRNA expression analysis was performed with the method described in Section 3.3.7. All expression results were normalised to miR-16 and relative quantisation was applied with the $2^{-\Delta\Delta Ct}$ method.

The miRNA expression results for Pre-miR-99a and Pre-miR-184 transfected 2102Ep cells are displayed in Figure 3.22. The results for Pre-miR-219 and Pre-miR-335 transfected cells are shown in Figure 3.23. MiR-10a is not expressed in undifferentiated NTera-2 cells and it is therefore not possible to measure relative expression of its expression in Pre-miR-10a transfected cells. This data is therefore not included in Figure 3.22. However, the raw Ct values of miR-10a in the Pre-miR-10a transfected NTera-2 cells is 21.73± 0.34, which is very similar to the raw Ct values of Pre-miR-184 transfected cells (Ct = 20.56± 0.19) and Pre-miR-99a transfected cells (Ct = 20.13±0.95). MiR-184 and miR-99a overexpression is very high in the respective controls, which strongly suggests a similarly high transfection efficiency for Pre-miR-10a.

In all functionally transfected samples, miRNA upregulation is very substantial and significant. The target miRNAs are overexpressed by a factor of approximately 400-31,000, although this probably depends largely on original miRNA expression. Transfection can be concluded to be successful due to this high amount and further analysis was conducted.

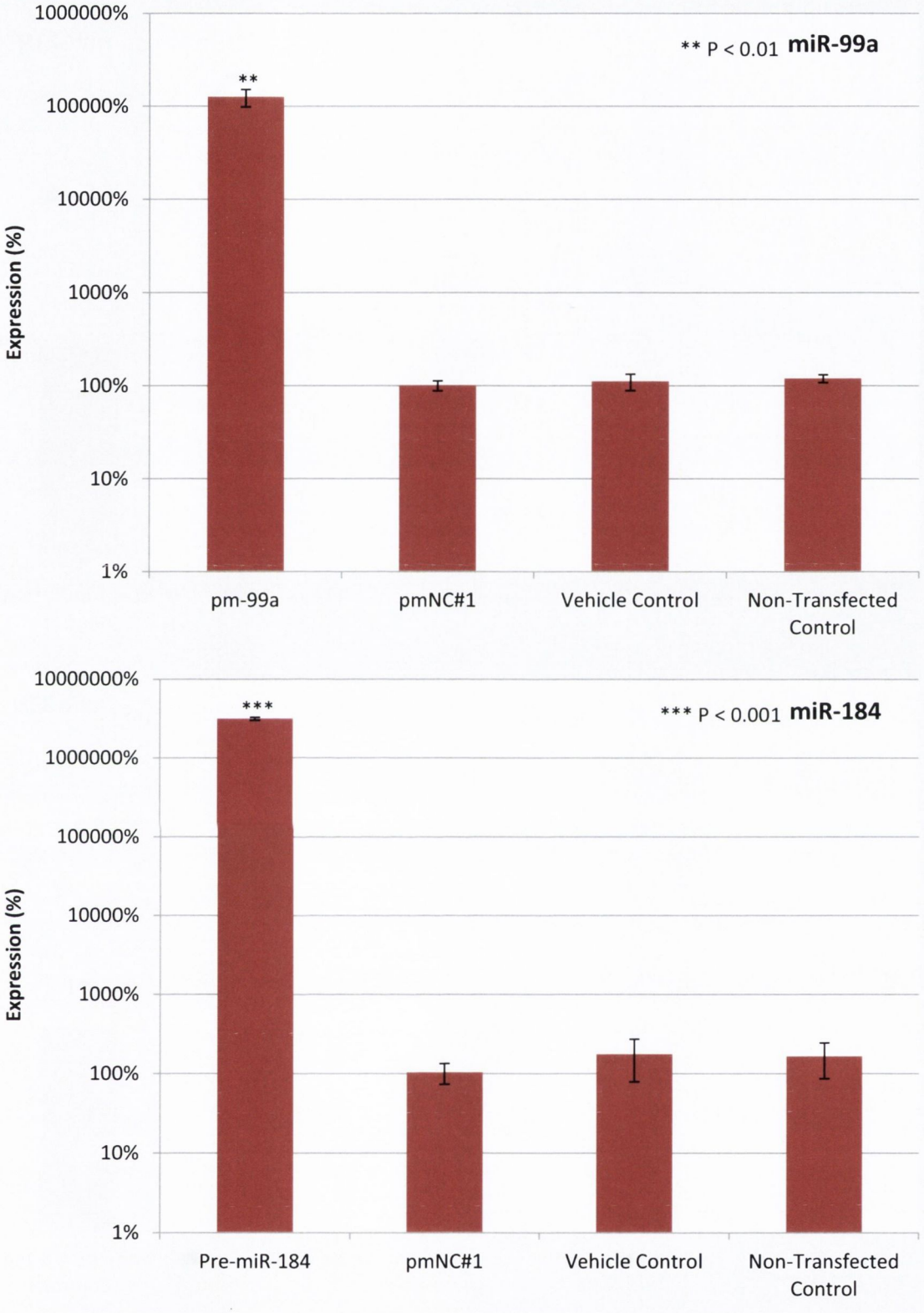


Figure 3.22—MiR-99a and miR-184 expression in functionally transfected NTera-2 cells. NTera-2 cells were transfected with Pre-miR™ miRNA precursor and incubated for three days. MiRNA expression, normalised to miR-16 expression is plotted on a logarithmic scale to account for high expression differences. Cells transfected with Pre-miR-99a showed a significant upregulation of miR-99a compared to its expression in pmNC#1-transfected cells ($P = 0.0031$), which was set at 100% expression. Cells transfected with Pre-miR-184 also showed a significant upregulation of miR-184 ($P < 0.0001$). No significant differences in miRNA expression were found among the pmNC#1-transfected cells and the vehicle and non-transfected controls. All experiments were performed in biological triplicate and the values shown is mean relative expression of these triplicates ± SD. To compare sample means, a t-test was used.

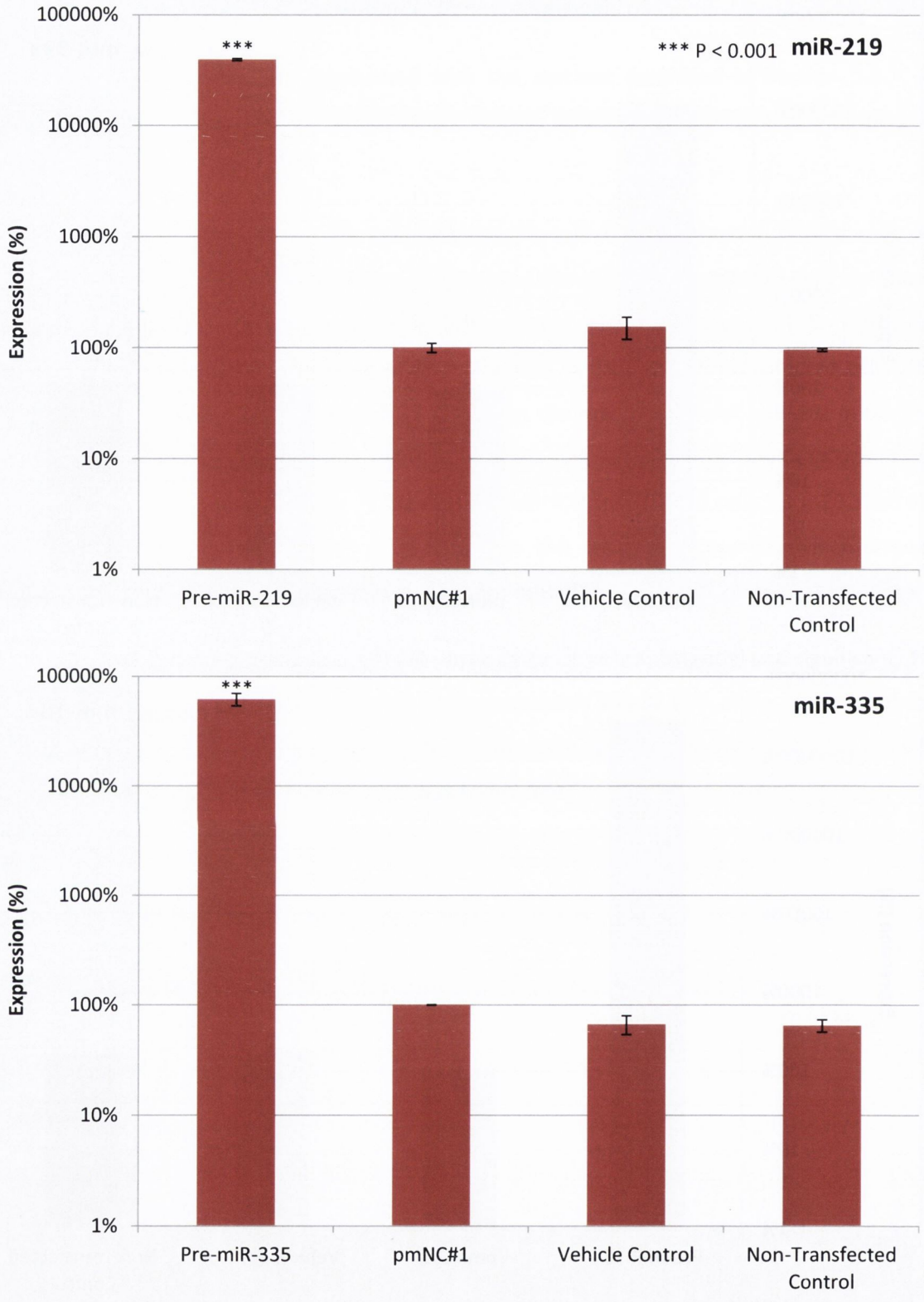


Figure 3.23—MiR-219 and miR-335 expression in functionally transfected NTera-2 cells. NTera-2 cells were transfected with Pre-miR™ miRNA precursor and incubated for three days. MiRNA expression, normalised to miR-16 expression is plotted on a logarithmic scale to account for high expression differences. NTera-2 transfected with Pre-miR-219 or Pre-miR-335 show a significant upregulation of these miRNAs compared to their expression in pmNC#1 samples ($P < 0.0001$), which was set at 100%. No significant differences exist among pmNC#1, vehicle and non-transfected controls. The expression in pmNC#1 samples was set at 100%. All experiments were performed in biological triplicate and the values shown is mean relative expression of these triplicates ± SD. To compare sample means, the t-test was used.

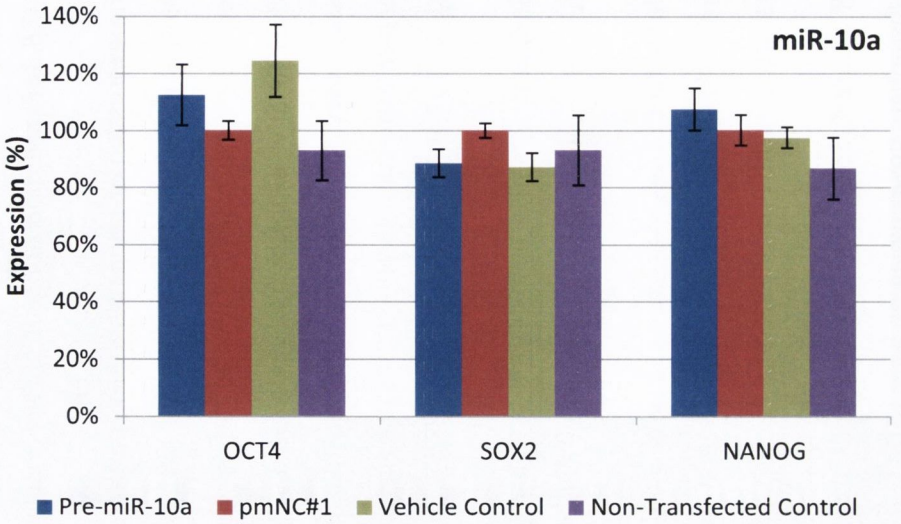
3.4.6.3- Expression Analysis of Pluripotency Markers

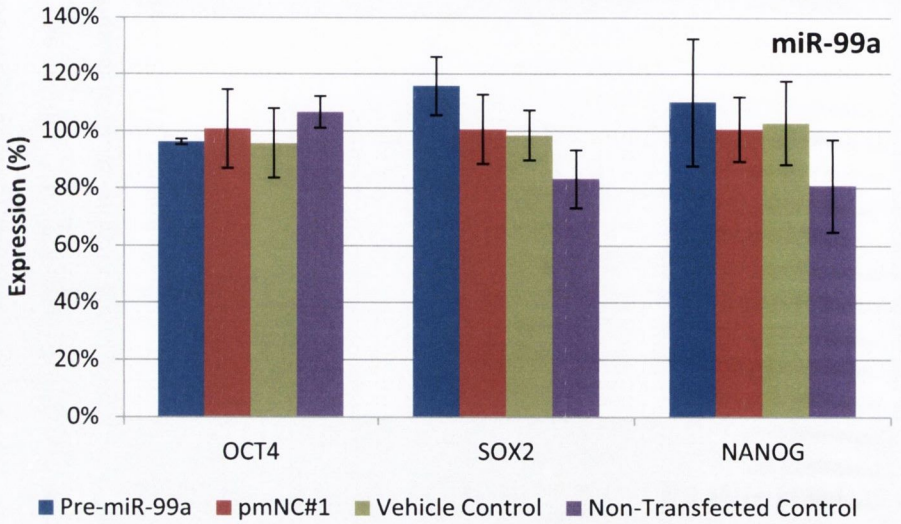
The expression of key pluripotency markers (OCT4, SOX2 and NANOG) was measured with the method in Section 3.3.7. The results were normalised to GAPDH expression and displayed as a percentage expression relative to the pmNC#1-transfected cells, which themselves were set at 100%. Relative expression was calculated using the $2^{-\Delta\Delta Ct}$ method. The TaqMan® Gene Expression assays used for measuring stemness marker expression are listed in Table 3.6. Pluripotency marker expression in NTera-2 cells transfected with Pre-miR-10a, -99a and -184 is shown in Figure 3.24.

Pluripotency marker expression in NTera-2 cells transfected with Pre-miR-10a and Pre-miR-99a showed no significant change with the negative controls.

Pluripotency were measured to assess whether a reduction in stemness was taking place in the Pre-miR-219 and Pre-miR-335 transfected cells. Pluripotency marker expression in NTera-2 cells transfected with Pre-miR-219 and Pre-miR-335 can be found in Figure 3.25.

Of all transfected cells, NTera-2 cells transfected with Pre-miR-335 show a significant and substantial downregulation of OCT4. This is accompanied with a almost significant downregulation of NANOG ($P = 0.0582$). Conversely, SOX2 is significantly upregulated. Overall, these results strongly suggest differentiation is taking place. The upregulation of SOX2 is an unusual feature, which cannot directly be explained with current results. Cells transfected with Pre-miR-219 showed no discernable change in pluripotency marker expression.





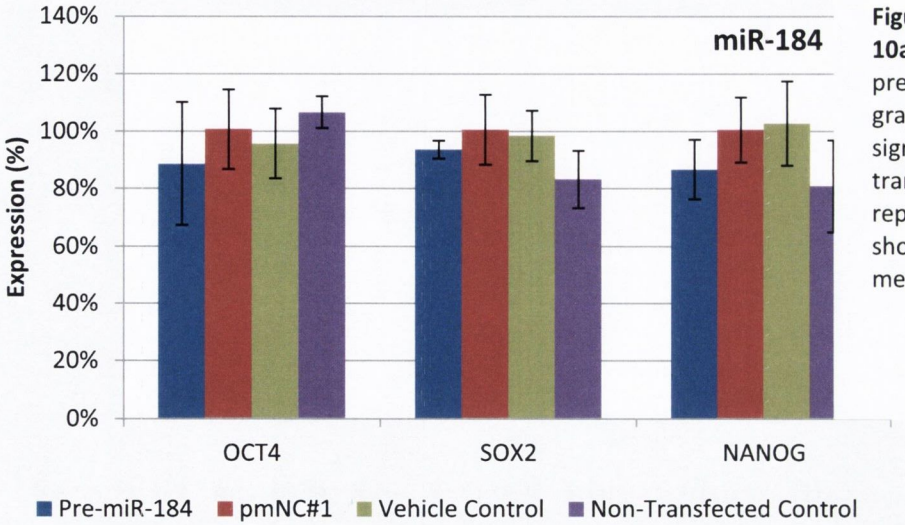


Figure 3.24—Pluripotency marker expression in NTera-2 transfected with Pre-miR-10a, -99a and 184. NTera-2 were transfected with 50 nM Pre-miR™ miRNA precursors and incubated for three days. Percentage gene expression shown in graph against that in pmNC#1-transfected cells, which was set at 100%. No significant alteration in pluripotency marker expression measured between cells transfected with Pre-miR-10a, -99a or -184 and their respective negative controls reported. All experiments were performed in biological triplicate and the values shown is mean relative expression of these triplicates ± SD. To compare sample means, a t-test was used.

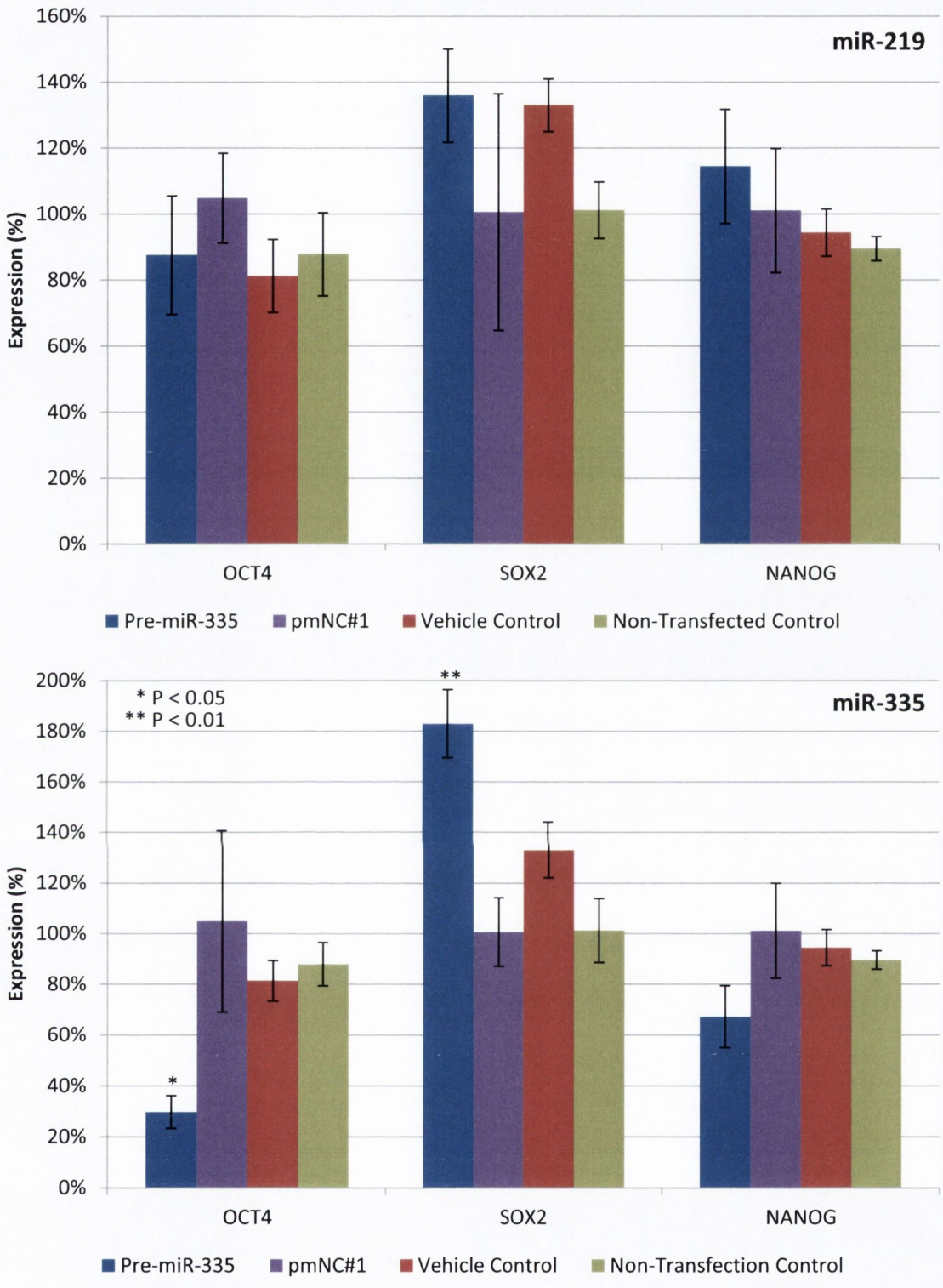
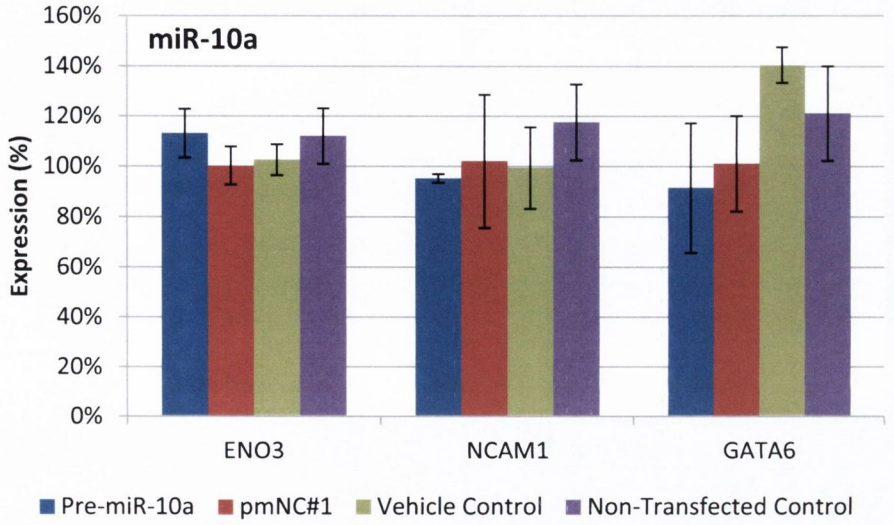
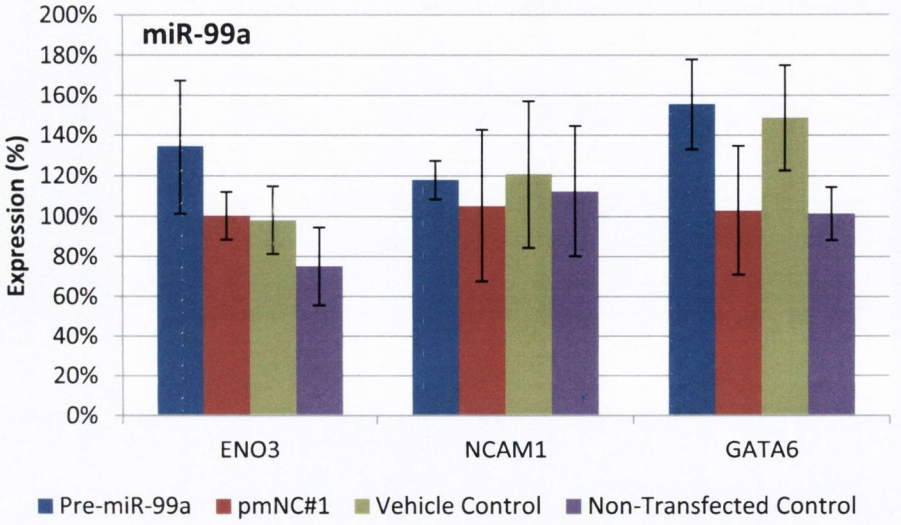


Figure 3.25—Pluripotency marker expression in NTera-2 cells transfected with Pre-miR-219 and Pre-miR-335. 2102Ep cells were transfected with 50 nM Pre-miR™ miRNA precursors and incubated for three days. Percentage gene expression is shown against the expression in pmNC#1 samples, which was set at 100%. No significant difference in expression between Pre-miR-219 and pmNC#1-transfected cells existed Pluripotency marker expression in NTera-2 cells transfected with Pre-miR-335 showed that OCT4 is significantly lower expressed in NTera-2 cells transfected with Pre-miR-335 compared to pmNC#1-transfected cells ($P = 0.0223$). SOX2 is significantly upregulated in NTera-2 cells transfected with Pre-miR-335 compared to pmNC#1-transfected cells ($P = 0.0017$). NANOG expression is marginally insignificant ($P = 0.0582$). All experiments were performed in biological triplicate and the values shown is mean relative expression of these triplicates ± SD. To compare sample means, a t-test was used.

3.4.6.4 — Expression Analysis of Differentiation Markers

Similarly to 2102Ep miRNA overexpression, differentiation marker (ENO3, NCAM1 and GATA6) expression was measured in NTera-2 cells transfected with Pre-miR-10a, -99a and -184. For this the gene expression protocol in Section 3.3.7 was used. TaqMan® Gene Expression assays used can be found in Table 3.6. Results are shown in Figure 3.26.





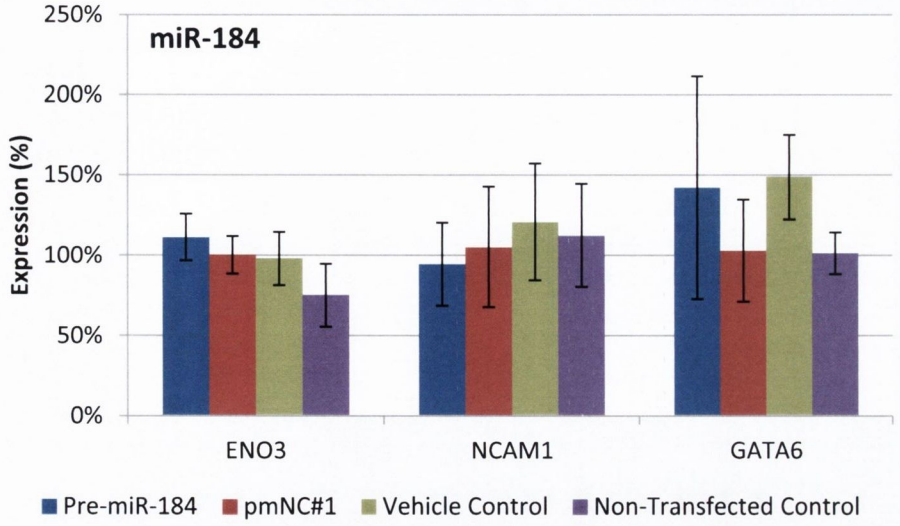


Figure 3.26—Differentiation marker expression in NTera-2 transfected with Pre-miR-10a, -99a and -184. NTera-2 were transfected with 50 nM Pre-miR™ miRNA precursors and incubated for three days. Percentage gene expression is shown against the expression in pmNC#1 samples, which was set at 100%. No significant alteration in pluripotency marker expression measured between cells transfected with Pre-miR-10a, -99a or -184 and pmNC#1-transfected cells could be determined. All experiments were performed in biological triplicate and the values shown is mean relative expression of these triplicates ± SD. To compare sample means, a t-test was used.

3.4.6.5 — Analysis of 6-day MiRNA Transfections in NTera-2 Cells

After viewing cell morphology at four days, it was decided to culture the cells to until 6-days post-transfection, at which the negative controls and Pre-miR-219 transfected cells were confluent. The cells were lysed on-plate and RNA was purified from a lysate fraction as described in Section 2.3.2.3. MiRNA and mRNA expression analysis was performed in exactly the same manner as was done in Section 3.4.6.2 and Section 3.4.6.3. Phenotypic change was observed during the extra 3-day incubation period and images were taken with a light microscope and displayed in Figure 3.27. At 6-days the Pre-miR-335 transfected NTera-2 cells still appeared flattened and enlarged. The pseudopodic projections became more frequent and pronounced, suggesting a possible progression to a neural fate, although this cannot be concluded from the images alone. Pre-miR-219 did not distinguish from the negative controls.

The expression of miR-335 and miR-219 as showed in Figure 3.28, revealed miRNA overexpression in the functionally transfected cells to still be significant and very substantial. At approximately 374-fold miR-335 and 214-fold miR-219 overexpression, the miRNA expression in the functionally transfected cells was about half in cells harvested at 6 days compared to those harvested at three days. This is expected, as the cells are split 1:2, half of the initially added Pre-miR™ miRNA Precursors would have remained in the wells post-split. This also indicates that the expression of these precursors are highly stable in NTera-2 cells.

OCT4 and NANOG were significantly downregulated by 53% and 44% respectively, as seen in Figure 3.29. OCT4 expression has slightly increased when compared to expression at three days. It is possible that OCT4 slightly re-expresses due to a reduction of functional Pre-miR-335 in the cells or a proliferation of cells that have avoided transfection. These latter reasons can be disputed Pre-miR-335 expression still remains very high due to the high stability of the Pre-miR™ miRNA Precursors. No strongly proliferating cells with an undifferentiated morphology was seen in the Pre-miR-335 transfected wells and as NANOG expression has fallen compared to the 3-day incubated cells, the presence of untransfected cells can probably be disputed. Differentiation marker expression was not tested, as the current results presented sufficient evidence that differentiation was taking place in Pre-miR-335 transfected NTera-2 cells and was not in Pre-miR-219 transfected cells.

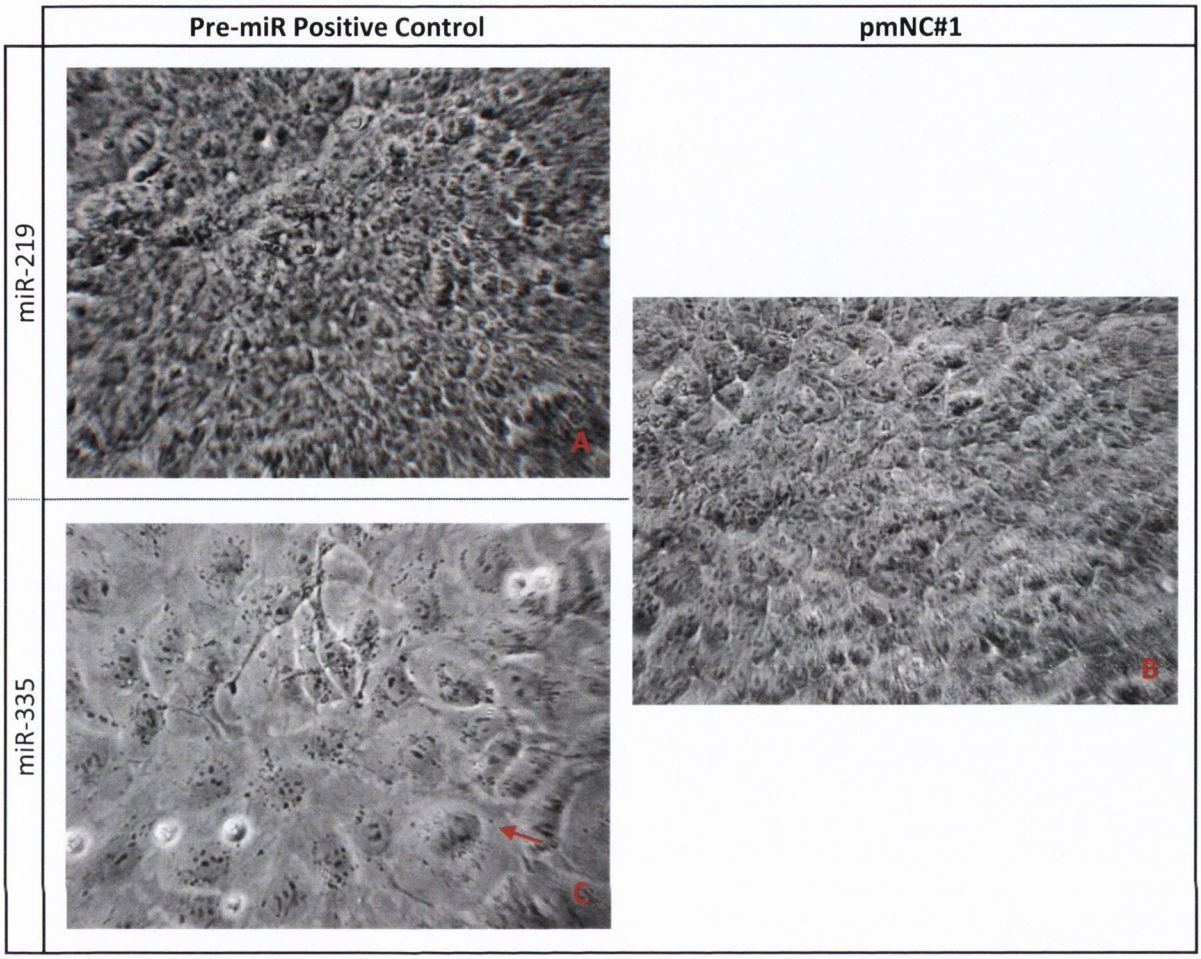


Figure 3.27—Morphology of NTera-2 cells transfected with Pre-miR™ precursors and incubated for 6 days. All images were photographed at a magnification of 320X. After 6 days of post-transfection incubation, cells transfected with Pre-miR-335 (Pre-miR Positive Control, C) showed a strong difference in morphology compared to compared to those transfected with pmNC#1 (B). As indicated by the red arrow, the cells had flattened, enlarged and less dense morphology, similar to that of Pre-miR-335-transfected cells after 3 days (see Figure 3.21). Similar to this time-point, transfected with Pre-miR-219 did not change (A). All images were photographed at a magnification of 320X.

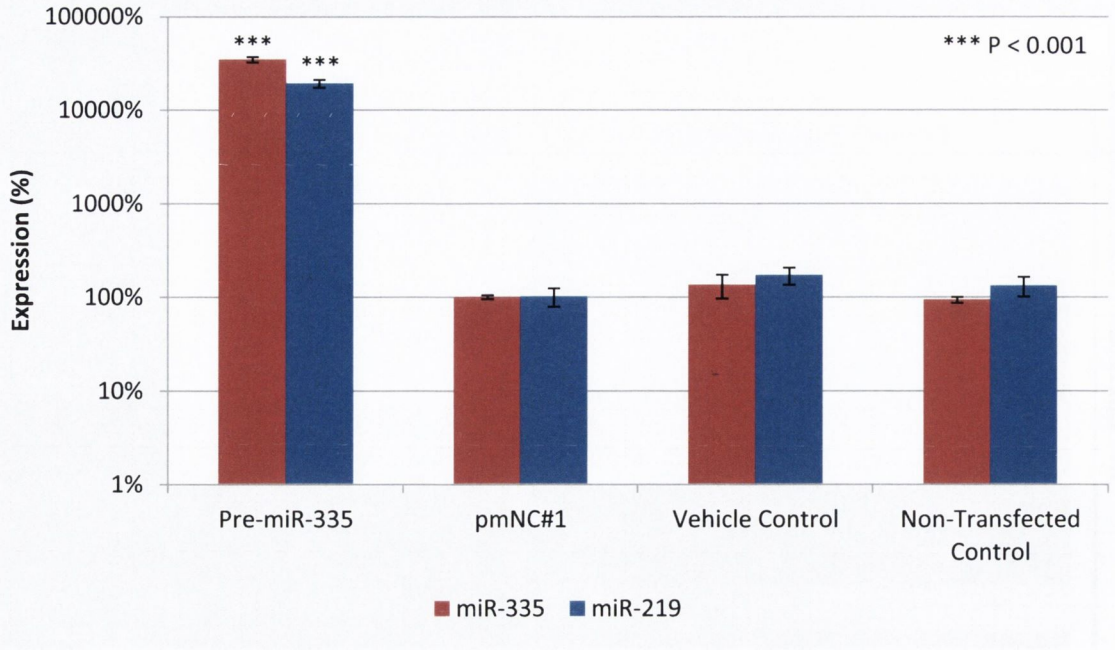


Figure 3.28—MiR-335 and miR-219 expression in NTera-2 cells transfected with Pre-miR-335 and Pre-miR-219 (6 days after transfection). NTera-2 cells were transfected with 50 nM Pre-miR™ miRNA precursors and incubated for 6 days. Percentage miRNA expression shown in graph against that in pmNC#1 samples, which were set at 100%. The expression of both miR-335 and miR-219 is significantly upregulated at day 6 post-transfection (Both: $P < 0.0001$). All experiments were performed in biological triplicate and the values shown is mean relative expression of these triplicates ± SD. To compare sample means, a t-test was used.

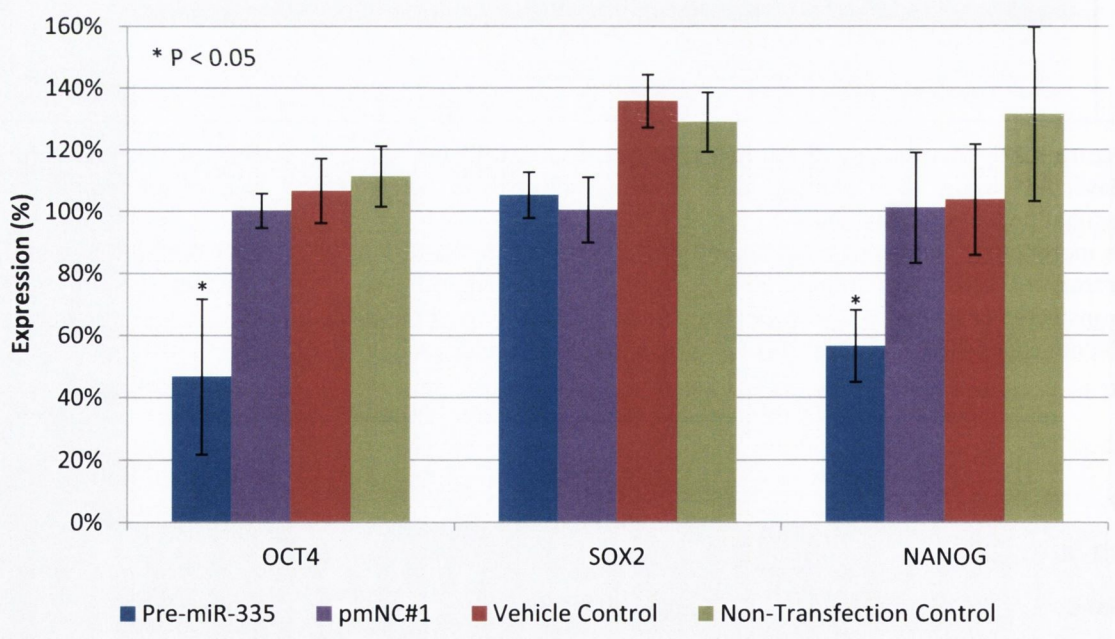


Figure 3.29—Pluripotency marker expression in NTera-2 cells transfected with Pre-miR-335 (6 days post-transfection). Both OCT4 and NANOG were significantly downregulated in Pre-miR-335 transfected NTera-2 cells compared with cells transfected with pmNC#1 ($P = 0.0224$ and $P = 0.0222$ respectively). SOX2 expression in Pre-miR-335-transfected cells does not significantly differ from those transfected with pmNC#1. ($P = 0.5551$), however SOX2 expression in both these controls differ from the vehicle and non-transfected controls ($P < 0.05$). The expression in pmNC#1 samples was set at 100%. All experiments were performed in biological triplicate and the values shown is mean relative expression of these triplicates ± SD. To compare sample means, a t-test was used.

3.4.6.6 — Protein Expression of Genes Potentially Targeted by miR-335

Western blotting is commonly used as an initial step of miRNA target validation. The targets of miR-335, suggested in Section 3.4.4, SOX2 and OCT4 were subjected to protein expression studies with Western blot. This was performed on both 3-day and 6-day post-incubation samples from previous experiments with a fraction of the cell lysate obtained with the miRVana™ PARIS kit. Protein obtained from NTera-2 cells incubated in 10^{-2} ATRA + hEC medium for three and 6 days alongside the functional transfection experiment was included as an extra control. Protein expression analysis was conducted as described in Section 3.3.8. The results are displayed in Figure 3.3.

OCT4 protein expression is visibly downregulated in NTera-2 cells transfected with Pre-miR-335 and in the cells treated with ATRA both at three and six days. This is also the case with SOX2 expression, although some downregulation is also detected in the pmNC#1-transfected cells. However this seems to recover at 6 days post-transfection. At three days post-transfection OCT4 and SOX2 expression in Pre-miR-335 transfected cells is seemingly lower compared to that in ATRA treated cells, whereas their expression is higher in Pre-miR-335 transfected cells compared to ATRA treated cells at 6 days post-transfection. This might support the proposal of direct targeting of SOX2 and OCT4 by miR-335 as the effect miR-335 has on these stemness markers might diminish over time, while ATRA continuously stimulates suppression of these markers. OCT4 protein expression seems to some extent reflect the its gene expression, which increases slightly at 6 days post-transfection compared to three days. Interestingly this inversely the case with SOX2 expression, which had a markedly higher expression at three days compared with the negative controls and showed no significant expression difference with these controls after 6 days. These results suggest a post-transcriptional regulation mechanism governing SOX2 expression, making miR-335 a potential component of this mechanism.

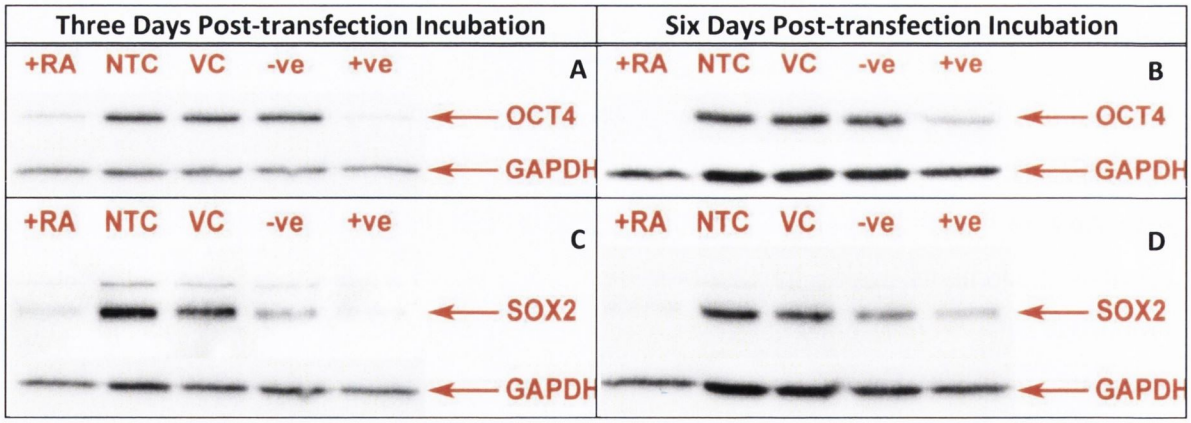


Figure 3.30—OCT4 and SOX2 protein expression in NTera-2 cells transfected with Pre-miR-335. After transfection with 50 nM Pre-miR-335, protein analysis was performed with the Western blot technique at three and 6 days post-transfection. This resulted in the silencing of OCT4 and SOX2 in Pre-miR-335 transfected samples (+ve). In the pmNC#1-transfected cells (-ve), the vehicle controls (VC) and non-transfected controls (NTC) SOX2 and OCT4 are still expressed. At 6 days OCT4 (B) and SOX2 (D) protein expression returns to some extent. Cells incubated in 10^{-2} ATRA (+RA) stop expressing OCT4 and SOX2 protein over 6 days. Western blots here are single replicates representative of the whole sample set.

3.5 — Discussion

3.5.1 — Successful Establishment of a Laboratory Protocol for Transient Ectopic MiRNA Overexpression in 2102Ep and NTera-2 Cell Lines.

The first aim of this chapter was to establish a system for the upregulation of miRNA in 2102Ep and NTera-2 ECCs. This system could then be used in the laboratory for future functional miRNA experiments and could also function as a basis for the transfection of other oligonucleotides into these ECC models. As mentioned, Pre-miR™ miRNA precursors were chosen for their previous performance in other cell lines in this laboratory, demonstrating a high and consistent transfection efficiency and stability. Recently Ambion® released a new generation of miRNA mimics, the miRVANA™ miRNA mimics, which have additional modifications and are single-stranded, in preference to the double-stranded Pre-miR™ precursors. This avoids any potential off-target mRNA inhibition otherwise a possibility with a Pre-miR™ precursor due to its minor strand. Unfortunately, these second generation miRNA mimics were not available at the time of this study.

Transfection optimisation was performed with Pre-miR-1 as an experimental control and pmNC#1 as a non-targeting control. MiR-1 targets the gene, PTK9 and knocks it down at the mRNA level, which was measured.[178,328] As found also seen in other studies with different cell lines, PTK9 mRNA expression was significantly and consistently reduced by more than 65% in 2102Ep and NTera-2 cells with relatively low cytotoxicity using the forward transfection method (see Section 3.4.1.2), indicating the successful optimisation of the ectopic miRNA overexpression protocol.[391–394]

3.5.2 — Functional Overexpression of MiR-335 and MiR-184 Induced Differentiation In ECCs

Two sets of miRNAs were selected based on different criteria. During a previous study performed in this laboratory, 2102Ep and NTera-2 cells were incubated in ATRA for three days and the differential expression of approximately 330 commonly detected miRNAs was measured.[274] From this data set miR-10a, miR-99a and miR-184 were selected because they were upregulated in both cell lines. It was reasoned that these miRNAs might be downstream targets the retinoid pathway. The core of this pathway consists of a conserved network of three nuclear retinoic acid receptors (RARs), which can bind retinoid compounds and regulate embryonal development by orchestrating the differentiation of ESCs.[150,395] They are also present in most ECC lines, including NTera-2 and 2102Ep cells.[390] It is thought that the inability of 2102Ep cells to respond to ATRA-

induced differentiation is because of an inhibition of crosstalk between the RAR and the Wnt/β-catenin networks.[390] RAR activation directly leads to the transcription of many different mRNAs and miRNAs and, thus, are directly regulated by retinoids, including, as now is known, miR-10a.[350] Publications at the time of this study and later research showed that miR-10a, miR-99a and miR-184 has tumour suppressive roles in cancer.[344,350,355,358,363,396] Furthermore, recent research has associated miR-10a with normal stem cell and CSC differentiation.[352–354]

The second set of miRNA, miR-219 and miR-335, were chosen on the basis of three criteria. Both these miRNA were among the highest upregulated miRNA in NTera-2 cells exposed to ATRA, which led to the supposition that they were potentially important in early ECC differentiation. Added to this, miR-219 and miR-335 were not changed in the differentiation-resistant 2102Ep cells after 3-day ATRA exposure, suggesting possible functions of miR-219 and miR-335 expression during ATRA-induced ECC differentiation. Finally, these miRNA were picked from a larger list of miRNA which predicted targets (TargetScan 5.1) were crucial hESC markers, as defined by a previous study.[335] Like the other miRNAs selected in this study, miR-335 also has reported tumour suppressive roles, although it can seemingly also function as an oncomir.[368,369] In addition, this miRNA has been characterised as a potent suppressor of tumour metastasis and has been shown to orchestrate the differentiation of mesenchymal stem cells (MSCs) and malignant glioma cells.[247,367,371–373,397] As with many miRNAs, it is possible that miR-335 coordinates developmental processes through a temporal and spatial coordination of its targets' expression. It has previously been established that miR-335 performs some of these carefully balance and timed tasks during osteogenic differentiation, epididymal development and during the development of breast cancer.[368,370,398] Finally, miR-219 has not been functionally linked to cancer, but is reportedly downregulated in some varieties of this disease.[399–402] It also has neural functions and roles in CNS development and differentiation, indicating that this miRNA is perhaps an early differentiation marker of neural precursors.[366,403,404]

The results of a 3-day incubation after transfection of the five miRNA in 2102Ep are shown in Section 3.4.5. In 2102Ep cells, all miRNA showed substantial upregulation. The induced upregulation of miR-184 showed limited differentiation in 2102Ep cells. These cells displayed flattened and enlarged morphology, which is typical of early ECC differentiation and could be indicative that this process was partially taking place. Of the stemness markers, SOX2 and NANOG showed significant downregulation (approx. 40%) and of the differentiation markers, GATA6 is upregulated by approximately 450%. Unusually, OCT4 expression levels remain unchanged, which was unexpected as this gene is regulated by SOX2 and NANOG and OCT4 is usually downregulated concomitantly in ECCs.[378] However, this might be the result of a delayed response of OCT4 or an

insufficient downregulation of the other two pluripotency factors. The induced knock-down of SOX2 in 2102Ep cells, described in Chapter 4, albeit by a higher percentage (approximately 10% remaining SOX2 expression), did result in the simultaneous downregulation of OCT4. As was shown in Figure 3.13, 2102Ep cells transfected with overexpressed miR-99a showed a morphology that resembled that of cells transfected with overexpressed miR-184. In addition to this, the differentiation marker, GATA6 was slightly upregulated (220%) and NCAM1 showed a marginally insignificant upregulation of 170% ($P = 0.0761$). However, none of the pluripotency markers were downregulated. These results were not sufficient to conclude differentiation was taking place in 2102Ep cells with induced miR-219 overexpression and other, undetermined cellular processes could be responsible for the altered cell morphology.

MiR-10a, -99a, -184, -219 and -335 were substantially upregulated in NTera-2 cells after transfection of their respective Pre-miR™ miRNA Precursors (results shown in Section 3.4.6). Of these miRNA, the cells with an overexpression of miR-335 showed a dramatic morphological change and a visible decline in cell number. Interestingly, their morphology was similar to NTera-2 cells that were subjected to three days of ATRA-induced differentiation. This might suggest differentiation was taking place in cells with overexpressed miR-335. This was supported by the substantial downregulation of OCT4 (approximately 30% expression remained) and the downregulation of NANOG by 35% after three days post-transfection. Interestingly, SOX2 expression increased by 80%. The expression of SOX2 during three days of ATRA-induced differentiation in NTera-2 cells normally remains relatively stable (see Section 3.4.6.3). To confirm these results, the post-transfection period was extended to 6-days. Although a slightly higher cell confluency could be observed after this incubation period, the morphology of NTera-2 cells with overexpressed miR-335 was still distinctly different than that of cells transfected with a non-targeting control, as can been seen in Figure 3.27. OCT4 expression was still low at approximately 40% remaining expression and NANOG expression decreased to below 60%. SOX2 regained its original expression level. These results led to believe that differentiation was taking place. The degree of this was different from that of NTera-2 cell exposed to ATRA, which almost completely lose OCT4 and NANOG expression after 6-days. This discrepancy could be explained by the possibility that a proportion of NTera-2 cells remained untransfected or insufficiently transfected with Pre-miR-335 and are left to form proliferating, undifferentiated subcolonies, whereas ATRA would affect all cells in the culture. Some regions of increased cell density were observed, however, immnostaining or immunosorting methods would have to be used to determine whether these colonies truly consist of undifferentiated cells. These methods were not available at the time of this study. Another explanation could be a potential compensatory role of SOX2, which was significantly higher expressed after three days post-transfection, potentially as a

reactionary response to the overexpression of miR-335. SOX2 regulates OCT4 and NANOG expression and therefore could have attempted to maintain OCT4 and NANOG expression in the inhibitory conditions of high miR-335 expression.

3.5.3 — OCT4 and SOX2 as Possible Targets of MiR-335

The pluripotency marker, OCT4, is a predicted and conserved target of miR-335 (TargetScan v5.1). With the miRanda target prediction software, SOX2 was predicted as an additional, but less conserved target.[322,340,341] Protein expression results, as measured by Western blot, showed a significant downregulation of OCT4 and SOX2 in NTera-2 cells for both 3- and 6-day time points after transfection of Pre-miR-335. Protein isolated from NTera-2 incubated in ATRA for three and sisx days was included in this experiment. The protein and mRNA expression of OCT4 was estimated to be relatively correlated over three and six days post-transfection. However, SOX2 mRNA expression at three days post-transfection was inversely correlated to the expression of its protein. Interestingly, cells exposed to ATRA also showed a downregulation of SOX2 protein, providing extra credence to the possibility that this gene is post-transcriptionally regulated in NTera-2 cells during ATRA-induced differentiation. This discrepancy has been recorded before in previous studies, but never discussed.[405]

The function of miR-335 in NTera-2 cells has not been elucidated in any previous known studies. However, functional publications of miR-335 exist, of which the most highly cited describe this miRNA as a tumour suppressor and suppressor of metastasis. The following paragraphs summarise the published functions of miR-335 and possible additional functions as hypothesised in this thesis based on the results shown in Chapter 3 and other published, but infrequently discussed data.

One of the first and most widely cited papers on miR-335 function directly linked this miRNA with a tumour suppressive function in metastatic breast cancer.[373] MiR-335 expression was lost in this disease causing the overexpression of its targets, SOX4, PTPRN2, TNC and MERTK, which are all genes important to tumour metastasis. Additional breast cancer suppressive properties of miR-335 were shown in a later study, which revealed that miR-335 cooperates with BRCA1 to regulate the expression of ESR1, IGF1R, SP1 and ID4.[368] In breast cancer, the loss of miR-335 expression has been shown to be due to epigenetic silencing and the reactivation of this miRNA can halt tumour reinitiation.[371] Further metastasis suppressive functions of miR-335 have been found in gastric cancer, which is mediated by the targeting of BCL2L2 and SP1.[367] Other tumour suppressive functions of miR-335 have been found in other cancers. One study showed the differential expression of miR-335 in many different cancer cell lines and showed that this miRNA could

regulate the cell cycle and stem the proliferation of an osteocarcinoma cell line by inhibiting p53 suppressor, RB1.[406] Other research has implicated miR-335 as an oncomir. One study showed that miR-335 can promote the proliferation and invasiveness of astrocytoma cells in cell culture and as xenografts by targeting the tumour suppressor, DAAM1.[369] MiR-335 has also been linked to carcinogenesis due to its rapid downregulation during early cell transformation.[407] Additional studies have correlated the loss of miR-335 with malignancy, metastasis and resistance.[408–411]

Other research suggests that miR-335 may have important functions during embryonal development and in adult stem cells. In adult stem cells, the transient regulation of miR-335 has been observed during the transitionary stages of adult stem cell establishment and differentiation. One study showed that upon differentiation of pre-adipocytes to adipocytes, miR-335 upregulated simultaneously with specific differentiation markers.[412] MiR-335 is upregulated in human MSCs (hMSCs) found in bone marrow, adipose tissue and articular cartilage and its overexpression in hMSCs has been shown to impair proliferation, migration and differentiation through the suppression of RUNX2.[372] In this study, it was furthermore shown that DKK1 decreased miR-335 expression through the Wnt signalling pathway.[372] Interestingly, DKK1 is also a validated target of miR-335.[370] Furthermore, IFNγ decreases miR-335 activity and Wnt3a increases it.[372] MiR-335 has also been associated with the promotion of osteogenic differentiation.[370] Some studies have shown the differential and transient expression of miR-335 during embryonal development and adult stem cell differentiation without further validating its function. This miRNA was upregulated during the differentiation of definitive endodermal cells to hepatocytes, but not from hESCs to definitive endodermal cells.[413]

MiR-335 is an intronic miRNA, so its transcription can be regulated by the promoter of its host gene, MEST. The correlated expression of MEST and miR-335 has been demonstrated.[372,414] This provides an opportunity to juxtapose mechanisms of miR-335 regulation in publications and datasets describing MEST. In addition to this, the regulation of miR-335 itself can be inferred.

Based on previous research and the results shown in this chapter, it can be hypothesised that miR-335 serves a dual function of guiding cell fate during embryonic development and regulating cell stemness and differentiation of adult stem cells. It can be hypothesised that this miR-335-mediated mechanism can respond to extracellular signalling and intracellular conditions to perform these functions.

The up or downregulation of miR-335 or its host gene, MEST, can be directly governed by upstream pathways activated or inhibited by extracellular signals. Examples of these are

activators or inhibitors of Wnt receptors, oestrogen receptors, LIF receptors and possibly many others.[368,372,415]

Although the precise regulation of MEST/miR-335 is still relatively unknown, MEST/miR-335 transcription is probably regulated by a set of activators and repressors of transcription that have important functions during embryonic development and in adult stem cells, but also in cancer. Myc, a cruicial stem cell transcription factor and downstream regulatory element of many important stemness pathways, has been found to be bound to the MEST/miR-335 promoter.[407] It has been speculated that Myc can repress MEST/miR-335 expression, although this has not been directly shown. LIF deprivation of mouse ESCs resulted in a decrease of STAT3 expression and a rapid increase of MEST expression.[415] INFγ can repress miR-335 expression.[372] The MEST/miR-335 promoter can potentially be bound by STAT3 and also by STAT1, both downstream regulatory elements of LIF and INFγ and important cytokines in stem cell maintenance.[372,415] Interestingly, STAT3 might also be a target of miR-335, although this has not been directly shown or validated.[372] In hMSCs, Chromatin Immunoprecipitation (ChIP) and computational analysis also have shown putative conserved binding sites in the MEST/miR-335 promoter for LEF1 and TCF4, which are downstream targets of the Wnt signalling pathway.[372] Putative binding sites for MAX and RXRα exist in the MEST/miR-335 promoter, suggesting MEST/miR-335 could be regulated by retinoids, as the previous data used for this thesis suggests.[274,372,416] It has also previously been shown that MEST/miR-335 expression is directly regulated by RARγ and RARα in F9 teratocarcinoma stem cells.[417,418] In myeloid leukaemia cells, the MEST/miR-335 promoter is bound by EGR1, which is, along with EGR2 and EGR3 potential targets of miR-335.[349,372]

The results in Section 3.4.6.3 show a significant increase in SOX2 mRNA expression and decrease in OCT4 and NANOG mRNA expression in NTera-2 cells three days after the induced overexpression of miR-335. OCT4 and SOX2 are predicted targets of miR-335 (TargetScan ver. 5.1 and miRanda), but have not yet been validated. When NTera-2 cells are induced to differentiate under ATRA conditions, OCT4 mRNA rapidly downregulates starting at approximately three days after induction.[375] In these cells, this has been shown to be caused by methylation of the OCT4 promoter site.[375] However, the loss of OCT4 mRNA and protein expression and the loss of SOX2 protein expression could possibly assisted by the direct suppression by miR-335, as the results in Section 3.4.6.6 indicate. Downregulation of NANOG would follow as this gene is regulated by both OCT4 and SOX2.[79] This could possibly be the mechanism of miR-335-induced differentiation of NTera-2 cells. However, unexpectedly, the deregulation of OCT4, SOX2 and NANOG mRNA has not been observed in 2102Ep cells upon miR-335 upregulation and, although OCT4 and SOX2 protein expression has not been measured in this cell line, the lack of visible differentiation of 2102Ep

cells seems to suggest, at first hand, that miR-335 has no significant function in the differentiation of these cells. Two explanations can be given for this. It is possible that miRNA-target suppression mechanisms or thermodynamic control of miRNA-target binding between miR-335 and OCT4 and SOX2 could be specific for NTera-2 cells and prevent miR-335-induced suppression in 2102Ep cells. Another, not mutually exclusive possibility is the existence of an intermediate element which can regulate the pluripotency markers and can be targeted by miR-335. SOX2 mRNA expression in NTera-2 cells increased by almost 2-fold upon the induced overexpression of miR-335. Although this could be an artefact, as SOX2 protein is decreased, it is important to note that an induced upregulation of SOX2 has before been shown to inhibit the transcription of OCT4:SOX2 target genes in ECCs and ESCs, including OCT4 and SOX2 themselves, but also NANOG and can also result the differentiation of ESCs.[419,420] Importantly, this effect can transpire even when the induction of SOX2 expression is below 2-fold. In addition to this, it has been shown that induced SOX2 overexpression can guide ESC differentiation towards a neuroectodermal fate at the expense of mesodermal and endodermal development.[421] Although more thorough investigation is needed, a possible hypothesis is that the overexpression of SOX2 mRNA, as mediated by miR-335, could cause the NTera-2 cells to differentiate. It is perhaps possible that miR-335 or a different post-transcriptional regulator inhibits SOX2 protein expression. One hypothesis is that temporal mechanisms could be operating to initially upregulate SOX2 after which a downregulation of all three core pluripotency markers follows. Besides a number of SOX2 promoter-binding transcription factors that are predicted targets of miR-335, SOX4, a validated target of miR-335 has been shown have the ability to activate SOX2 expression in glioma CSCs.[422]

Although it is unknown whether a commitment to a particular cell fate is induced by miR-335 overexpression in NTera-2 cells, an participation of miR-335 in ATRA-induced neuroectodermal differentiation of NTera-2 cells could be predicated from aforementioned established theories and the findings in this chapter.

Phenotypically, 2102Ep cells are said to resemble hESCs quite closely when in the undifferentiated state, whereas undifferentiated NTera-2 cells, although still pluripotent and relatively uncommitted, present a phenotype that partially resembles hESCs but also cells of the early germ layer.[280,317,318] Although only few publications mention miR-335 during hESC differentiation and embryonal development, it can possibly be said that, as this miRNA is undetected in most undifferentiated ESCs, but more commonly in early embryonic dermal cells as well as later progenitor cells, miR-335 is an active player in ESC commitment and early tissue development. The regulation or attenuation of developmental and stemness genes by miR-335 could be transient and local. For example, miR-335 is only expressed during a specific time point during

mouse neural tube development.[423] Another example is the increased expression of miR-335 during mouse and human lung development, after which expression gradually reduces during post-natal maturation of mice.[424] A similar post-natal pattern is seen in rat epidymal development and p21 regulator, RASA1 was proposed as a miR-335 target of importance in this process.[398]

Although a number of publications associate miR-335 with embryonic development, adult stem cell maintenance and differentiation and embryonic and adult tissue (re)generation, this miRNA has not been properly reviewed in relation to these functions. Also, its tumour suppressive and anti-metastatic properties have not yet been reconciled with these aforementioned functions. It is quite possible that these two *modi operandi* of miR-335 are greatly intertwined and future research , including more comprehensive reviewing, could reveal the precise local and temporal functions in normal and cancerous cells and tissues.

MiR-184 has been previously shown to inhibit neuroblastoma growth by targeting AKT2, however no reduction of this gene's expression was seen in NTera-2 cells exposed to ATRA for three days.[362,363] This does not mean that AKT2 is not targeted, as miR-184 could merely inhibit the translation of its mRNA. However, one notable potential target that was downregulated at the mRNA level in NTera-2 cells, is TCF7L1. This gene is a potential target of miR-184 as predicted by five different target prediction software tools. This downstream transcription factor of the Wnt (activating) and TGFβ (repressing) pathway cooperates with LEF transcription factors to regulate gene expression during embryonal development and stem cell differentiation.[425] TCF7L1 appears to be able facilitate or inhibit differentiation depending on cell type.[426,427] Optimisation of TCF7L1 antibodies available at the time was attempted, to determine TCF7L1 protein downregulation. However, although some preliminary screening might suggest that TCF7L1 protein is downregulated after miR-184 overexpression, antibody binding specificity was not sufficient to meet experimental standards of certainty. Further experimentation with more reliable available antibodies might reveal this gene to be a target of miR-184 and might elucidate the differences between miR-184 function in 2102Ep and NTera-2 cells.

In summary:

- The overexpression of miR-335 resulted in the differentiation of NTera-2 cells and the strong downregulation of OCT4 and NANOG transcripts and the upregulation of SOX2.
- The overexpression of miR-335 resulted in the downregulation of OCT4 and SOX2 protein. As these are its predicted targets, this could be directly caused by miR-335
- The overexpression of miR-184, to a limited extent, resulted in the the downregulation of SOX2 and NANOG in 2102Ep cells, but did not visibly differentiate these cells.

CHAPTER FOUR

INDUCED SILENCING OF GENES IN 2102Ep AND NTera-2

4 — Induced Silencing of Genes in 2102Ep and NTera-2 Cells

4.1 — Introduction

Due to their apparent similarity, the targeting of CSCs, without affecting the homeostatic function of normal stem cell is a major challenge for drug development.[158] This challenge is exacerbated by the difficulty of detecting rare CSCs and the irregular nature of their phenotype. Additionally, any potential CSC-specific drug targets are likely to vary between cancer types as the treatment effectiveness can also depend on the primary or secondary host tissues and the anatomy of the cancer.[158] In addition, drug treatment could also require modifications to account for the malignant stage of the disease, as the distinct gene expression of the contributing CSCs could dynamically alter between stages.

So far, a handful of genes have been identified in various CSC cell lines, which have the potential to be selective drug targets, without affecting their normal stem cells counterparts *in vivo*.[428–431] One interesting study showed that the embryonal carcinoma cell lines, NTera-2, NCCIT and P19 cells, were more sensitive to differentiation due to NANOG downregulation induced by the histone deacytelase inhibitor, apicidin, compared to ESCs.[290] Because of the reduced malignancy, attributable to reduced potency and self-renewal, this compound might therefore serve as a drug with a specificity to the embryonal carcinoma phenotype. Another therapeutic option is the specific targeting of genes which are functionally mutated in CSCs, contributing to an altered mechanistic function, which otherwise would be absent in normal stem cells. For example, attractive targets would be fusion genes (the genetic fusion of two or more gene loci, facilitated by mutations and resulting in a single transcript and protein) with important CSC regulatory functions and are not endogenous to normal cells.[432] As briefly discussed in Section 3.1, an alternate strategy for CSC gene therapy could encompass a targeted, drug delivery based approach rather than targeting CSC-specific genes. This would require accurate profiling of CSC-specific immnunomarkers, which could guide CSC marker-specific vessels (e.g. antibody-conjugated nanoparticles) containing siRNA or specific (ant)agonists which target stem cell genes.[312]

In Section 1.4 and Section 3.1, the aspects of CSC differentiation therapy are discussed. In these Sections, a number specific genes and pathways are specified which have been and could be involved in the maintenance of the stem cell phenotype. The Wnt/β-catenin, Hedgehog, TGFβ-related pathways were given as examples of targets for differentiation treatment and have been investigated for this purpose. Similarly to Chapter 3, CSC differentiation was the focus of this chapter.

The RAR pathway and its involvement with ESC and ECC differentiation has been well characterised. RARs are nuclear receptors consisting of heterodimers of one of three RAR proteins (RARα, RARβ and RARγ) combined with one of three RXR proteins (RXRα, RXRβ and RXRγ).[433] Isoforms exist of RAR and RXR proteins, resulting in a diverse combination of RAR heterodimers.[433] However, in some cases two RXR proteins can form a homodimeric RAR.[434] Upon activation, these can bind to gene promoter sequences called RAR elements (RAREs) and retinoid X response elements (RXREs).[434] Different types of retinoid compounds can bind to different RAR types, resulting in the transcription or repression of specific genes.[434]

The study conducted in this chapter was based on the results of a previous experiment performed by this laboratory. Here, 2102Ep and NTera-2 cells were differentiated by incubation in ATRA for three days. Whole-genome analysis was performed and the differential gene expression between experimental samples and control samples was measured. ATRA agonise RARs upon which they bind to specific RARE motifs in the promoter regions of many genes, particularly ones involved in embryonic development, such as the Homeobox (HOX) genes and other RARs.[434] A large number of genes containing RAREs and RXREs in their promoter regions were significantly deregulated in both 2102Ep and NTera-2 cells after ATRA-induced differentiation. From the list of downregulated genes in NTera-2 cells previously generated during the aforementioned study, five were picked for the study in this chapter with the aim to induce knock-down in NTera-2 and 2102Ep cells. These are listed in Table 4.9, on page 189 further on in Section 4.4.4, and were among the most downregulated genes in NTera-2 cells during early differentiation. An important goal was to target genes that did not have a previously reported function in hES differentiation, as these cells were considered to be the NSC counterpart of hECCs (see Section 1.13 for further information on this). To filter out genes associated with hESC differentiation a number of measures were taken, which are described in Section 4.4.4. The targeted genes selected for knock down were DPPA3, SHISA2, RBM47, AK3L1 and VSNL1. It was believed that by silencing these genes in NTera-2 and 2102Ep cells, it might push these cells to differentiate. DPPA3, also called Stella, was a particularly interesting potential target, as previous studies revealed that this gene is expressed in both 2102Ep and NTera-2 cells, while unexpressed in hESCs, signifying a potentially important functional difference between these cell types.[280,318]

In addition to silencing the five genes described in the previous paragraph, the decision was made to silence SOX2 in both cell lines. This gene is stably expressed during ATRA induced differentiation of NTera-2 cells. The cell fate of NTera-2 cells upon their differentiation by retinoids is skewed towards cells that constitute the CNS, which require SOX2 for neural homeostasis (see Chapter 3 and Chapter 5 for more details).[375,376] Because of the role of SOX2 at

maintaining pluripotency in hESCs, the hypothesis was held that the downregulation of this gene could potentially induce differentiation in 2102Ep and NTera-2 cells.

Transient gene silencing in this chapter was performed with exogenous siRNA, which utilises cells' endogenous RNAi machinery to selectively and transiently suppress mRNA translation by inducing the degradation of these molecules. This is the main function of endogenous siRNAs to specifically control a large group of genes and pathways, as well as foreign RNA sequences, such as viruses.[435] As with miRNA, siRNA-mediated transcriptional silencing is performed by RISC.[435] SiRNA preferentially load into the Argonaute protein, AGO2, and forms, together with other RISC components, 'siRISC'.[173] This loaded RISC is responsible for the suppression of genes by the targeted cleavage of mRNAs, which has a greater impact on protein expression than the more subtle and temporary mRNA repression as commonly induced by miRNA (see Section 1.6 for a more detailed description). The discovery of the RNAi mechanism was first made in 1998, by the transient delivery of exogenous double-stranded RNA sequences into *C. elegans* cells.[436] The use of exogenous siRNAs for transient gene suppression exists now as one of the most widely-used biomolecular laboratory techniques and is also a widely explored method for the treatment of various diseases, including cancer.[437] Exogenous siRNAs can be introduced into cells in two forms. For a transient suppression of protein expression, double-stranded RNA can be delivered by various methods, which are subsequently cleaved to mature siRNAs. For a long-term solution for gene knock-down over many mitotic cycles, vector sequences containing short hairpin RNAs (shRNAs), precursors to siRNAs, can be transfected or virally delivered into cells, which are subsequently spliced into the genome by the cell's endogenous machinery.[437] From this, shRNAs, which structurally resemble pre-miRNAs, are transcribed, exported out of the nucleus and subsequently processed by Dicer into mature siRNAs.[171,437] For the suppression of specific genes in this chapter, double-stranded siRNA precursors (first method) were used, as permanent gene downregulation was not required experimentally. A perfect complementarity of siRNAs to their target genes is required for successful mRNA cleavage and degradation.[435] In this chapter, perfect match exogenous siRNAs, simply abbreviated as 'siRNAs' for this chapter, are used to transiently silence specific gene expression as is described in Sections 4.3.2.1 and 4.3.3.1. This was chosen over the use of shRNAs, as only short term gene knock-down was required for the designed experiments. Furthermore, siRNA experiments are easier to perform and more commonly used, providing high-quality results from extensive experimental and product design optimisation.

4.2 — Chapter Aims

The aim of this chapter was to induce differentiation in 2102Ep or NTera-2 cells by transiently silencing the expression of the pluripotency marker, SOX2 and also five highly expressed genes which were strongly downregulated in NTera-2 cells during early differentiation and believed to potentially contribute to CSC differentiation.

4.3 — Materials and Methods

In this subchapter, methods supplemental to the general methods described in chapter 2 is provided. These methods are exclusive to this chapter and the referenced Sections in this subchapter supersede the methods described in chapter 2. If no supplemental methods are given, the general methods are referenced to.

4.3.1 — Cell Culture

2102Ep and NTera-2 cells were maintained as detailed in Section 2.1.5 and kept under 50 passages. Cells were counted as described in Section Section 2.1.7 and cells were plated at specific densities on 6-well cell culture plates. All incubations were performed in a humidified environment at 37°C and 5% CO_2.

4.3.2 — Optimisation of RNAi

For this study gene expression was silenced with Ambion's Silencer™ Select siRNAs *(4427037, Ambion, Applied Biosystems, Life Technologies)* were used. SiRNAs are short double-stranded oligonucleotides and are therefore of similar chemical compositionas the Pre-miR™ miRNA Precursors used in Chapter 3. Because of this, transfection conditions similar to what was used for Pre-miR™ miRNA Precursors were used for the transfection of siRNA into 2102Ep and NTera-2 cells.

4.3.2.1 — Transfection of siRNA

As with Pre-miR™ miRNA Precursor forward transfection optimisation, transfections were performed in triplicate on 96-well plates with the forward transfection method. Initial transfection optimisation was performed with different siRNA concentrations, but transfection agent concentrations were kept constant at the final volumes established in Section 3.3.5.2. These volumes are listed in Table 4.1. Ambion recommends a final transfection concentration range of 5-30 nM for their Silencer™ Select siRNAs. SOX2 was the first functional target chosen for silencing in NTera-2 and 2102Ep cells. The siRNA used to target this gene was therefore used for optimising transfection, with the aim of applying the resultant protocol to further siRNA transfections. For SOX2 silencing the Silencer™ Select siRNA was transfected at final concentrations specific to each cell line as listed in

Table 4.2. In addition to SOX2, transfection optimisation in NTera-2 cells was also performed with GAPDH siRNA at concentrations listed in

Table 4.2. For siRNA transfection in 2102Ep and NTera-2 cells, cells were harvested and seeded in sterile 96-well cell culture plates *(167008, Nunc, Thermo Fisher Scientific)* at 6000 cells/well in 100µl hEC medium 24 hours before transfection. Transfection master mixes were prepared as per manufacturers' instructions by preparing 20µl/well transfection master mixes which consisted of serum-free Opti-MEM® I medium containing transfection agents and siRNA diluted from a 5 µM stock solution to the required concentrations. To let the transfection complexes to form, the transfection master mixes were incubated for 15 minutes at room temperature. For Lipofectamine™ 2000 a five minute preincubation step of only transfection agent in Opti-MEM® I medium was included. During incubation, the old cell medium was discarded and 80 µl Opti-MEM® I (RT°) was added after a wash with 1X PBS (37°). The cells were overlayed with 20 µl transfection master mix and the plate was manually shaken briefly. Non-transfected controls were included by replacing old cell medium with 100 µl fresh hEC medium (37°). The plates were incubated in a humidified environment at 37° and 5% CO_2 for 6 hours after which the media were replaced with 100 µl hEC medium (37°). The cells were further incubated for a defined period. For protein expression these transfections were repeated on 6-well plates. The volumes were scaled to this plate size accordingly.

Table 4.1—Volumes of transfection agent used for transfection optimisation and functional siRNA transfections in 2102Ep and NTera-2 cells.

Cells	Transfection Agent	Volume/well	
		96-well plates	6-well plates
2102Ep	Lipofectamine™ 2000	High (0.375 µl)	High (10.57 µl)
NTera-2	Lipofectamine™ RNAiMAX	Medium (0.2 µl) and High (0.3 µl)	High (8.46 µl)

Table 4.2—Different final concentrations of siRNA used for transfection optimisation in 2102Ep and NTera-2 cells.

Cell Line	siRNA	Final Conc.	Product ID
2102Ep	siSOX2	10 nM 15 nM	s13295
NTera-2	siSOX2	5 nM 10 nM 15 nM	s13295
	siGAPDH	30 nM	s5573

For NTera-2 cells, the Gene Expression Cells-to-Ct kit was used to extract RNA and perform cDNA synthesis and qRT-PCR. The protocols for this are described in Section 2.3.2.1 and Section 2.5.2.2. For 2102Ep cells, RNA was purified with the protocol explained in Section 2.3.2.4. cDNA synthesis and qRT-PCR were performed as detailed in Section 2.5.2.1. Raw expression data was processed

with the $2^{\Delta\Delta Ct}$ method. Transfection efficiencies were shown in a percentage of remaining expression of the target gene.

4.3.2.2 — Gene Expression Analysis

RNA purification from 2102Ep cells was performed with the MagMax™-96 Total RNA Isolation kit as described in 2.3.2.4. Subsequent DNA removal was performed by the method detailed in Section 2.3.2.5. After analysis of RNA quantity and purity per the method in Section 2.4.2.1, gene expression analysis was performed as described in 2.5.2.1. RNA isolation and gene expression analysis in NTera-2 cells was performed with the Cells-to-Ct™ Gene Expression kit as described in Section 2.3.2.1 and Section 2.5.2.2. The TaqMan® primer/probe assays used for gene analysis during transfection optimisation are listed in Table 4.3.

Table 4.3—TaqMan® primer/probe assays used during siRNA transfection optimisation.

Gene	Product ID
SOX2	Hs01895061_u1
GAPDH	Hs01053049_s1
B2M	Hs04260366_g1

4.3.2.3 — Protein Expression Analysis

Protein lysates were obtained with the *miR*VANA™ PARIS kit *(AM1556, Ambion, Applied Biosystems, Life Technologies)* as described in Section 2.6.2.1. For Western blotting the protocol was used after measuring and equalising protein concentrations using the protocol from Section 2.6.2.2. All samples were run on a single 12% gel as described in Section 2.6.2.3. Protein transfer to the membranes was performed as described in Section 2.6.2.4. Membranes were blocked in blocking buffer for 1 hour at room temperature and subsequently incubated overnight at 4°C in primary antibodies as described in Section 2.6.2.5. After washing and secondary antibody binding, ECL was used for visualisation of target proteins after a second wash using the protocol from Section 2.6.2.7. For the semi-quantitative normalisation of sample concentrations, probing was repeated as described above after membrane stripping as described in Section 2.6.2.8. For all primary antibodies used, including their concentrations, see Table 4.4. For secondary antibodies, see Table 4.5.

Table 4.4—Primary antibodies for Western Blot.

Gene	Host	Manufacturer	Product Code	Dilution Used
GAPDH	Mouse	Abcam plc.	ab8245	1:25,000
B2M	Mouse	Abcam plc.	ab759	1:1000
SOX2	Mouse	Abcam plc.	ab75485	1:500

Table 4.5—Secondary antibodies for Western Blot.

Antigen	Host	Manufacturer	Product Code	Dilution Used
Anti-mouse	Mouse	Jackson Immunoresearch	115-165-062	1:5000
Anti-rabbit	Donkey	Santa Cruz Biotechnology	sc-2313	1:2000

4.3.3 — Functional Gene Silencing

4.3.3.1 — Functional Transfection of siRNA

For gene and protein expression studies, functional RNAi was performed with Silencer™ Select siRNA *(4427037, Ambion, Applied Biosystems, Life Technologies)* in 6-well plates. 2102Ep and NTera-2 cells were transfected with the forward transfection method described in Section 4.3.2.1. NTera-2 and 2102Ep cells were transfected with transfection agent concentrations shown in Table 4.1. RNAi in 2102Ep and NTera-2 cells was performed with 10 nM and 30 nM siRNAs respectively from 5 μM stock dilutions. The siRNAs used are listed in Table 4.6. Plate layouts for SOX2 silencing in 2102Ep and NTera-2 cells, and the silencing of five genes in both cell lines are shown in Figure 4.1. Post-transfection incubations as defined were maintained in humidified conditions at 37°C and 5% CO_2, while refreshing the hEC medium every 24 hours. For the five gene knock-down experiment all functional siRNA and the ssNC#1 were combined on Plate 1. The vehicle, non-transfected control and differentiation control were cultured on Plate 2. The latter control consisted either of NTera-2 cells exposed to 10^{-2} μM ATRA in hEC medium for the duration of the post-transfection incubation time or of 2102Ep cells transfected with siSOX2 in a manner which was consistent with the experiment.

Table 4.6—Product IDs and sequences of functional Silencer™ Select siRNA used in Chapter 4.

siRNA	Product ID
siDPPA3	s51495
siSHISA2	s51997
siRBM47	s230615
siAK3L1	s1213
siVSNL1	s14826
siSOX2	s13295

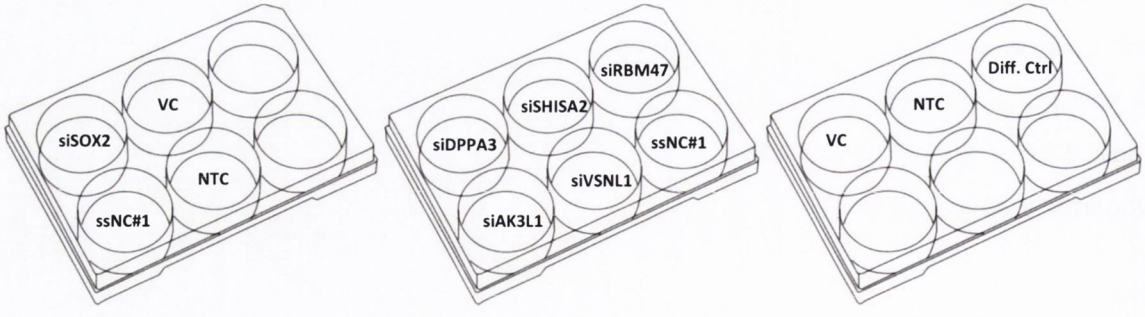


Figure 4.1—Cell culture plate layout of functional transfections of siRNAs. The siSOX2 transfections were performed on a single plate together with the accompanying negative controls; ssNC#1, vehicle control (VC) and non-transfected control (NTC). For the knock-down of five genes, functional siRNAs were transfect into cells on single plate, together with ssNC#1. The vehicle and non-transfected controls were maintained on a separate plate, together with a differentiation control. This image represents a single biological replicate.

4.3.3.2- Gene Expression Analysis

In both cell lines RNA purification was performed with the *miR*VANA™ PARIS kit as described in Section 2.3.2.3. After assess RNA purity and quantity with the method in Section 2.4.2.1, gene expression analysis was performed with the method in Section 2.5.2.1. The TaqMan® primer/probe assays used for gene expression analysis during functional transfections are listed in Table 4.7

Table 4.7—TaqMan® primer/probe assays used during functional gene silencing.

Gene	Product ID
DPPA3	Hs01931905_g1
SHISA2	Hs01590823_m1
RBM47	Hs00219308_m1
AK3L1	Hs03405743_g1
VSNL1	Hs02386966_m1
SOX2	Hs00602736_s1
OCT4	Hs01654807_s1
NANOG	Hs02387400_g1
ENO3	Hs00266551_m1
NCAM1	Hs00169851_m1
GATA6	Hs00232018_m1
GAPDH	4331182

4.3.3.3 — Protein Expression Analysis

Protein expression was performed with SDS-PAGE and Western Blot using similar methods as previously described in methods described in Section 4.3.2.3. Protein transfer was performed as described in Section 2.6.2.4 for either for 75 minutes at 100V or overnight at 45V and 4°C. Table 4.8 shows the primary (1st AB ID) and secondary antibodies (2nd AB ID) used in this chapter. Many showed inconclusive results. Some primary antibodies were blocked with inhibition peptides to determine the correct protein band. This is also shown in Table 4.8. The protocol for this can be found in Section 2.6.2.6.

Table 4.8—Primary and secondary antibodies for Western blot.

Gene	Predicted MW	1st AB ID	Clone Type	Host	Dilution	2nd AB ID	Dilution	Inhibition Peptide
DPPA3	17.9 kDa	sc-51435	Poly	Goat	1:200	sc-2022	1:2000	sc-51435-P
		sc-67250	Poly	Rabbit	1:200	sc-2313	1:2000	
SHISA2	31.4 kDa	sc-84640	Poly	Rabbit	1:200	sc-2313	1:2000	sc-84640-P
		sc-84642	Poly	Rabbit	1:200	sc-2313	1:2000	
		ab107724	Poly	Rabbit	1:200	sc-2313	1:2000	
RBM47	**64.1** or 56.9 kDa	ab94638	Poly	Rabbit	1:100	sc-2313	1:2000	
AK3L1	25.6 kDa	ab97382	Poly	Rabbit	1:1000	sc-2313	1:2000	
VSNL1	22.1 kDa	ab58053	Mono	Mouse	1:100	sc-2314	1:2000	
		ab102476	Poly	Rabbit	1:500	sc-2313	1:2000	
GAPDH	36.1 kDa	ab8245	Mono	Mouse	1:25,000	115-165-062	1:5000	

4.4 — Results

This subchapter contains the results of siRNA transfection optimisation and functional gene silencing.

4.4.1 — Optimisation of RNAi

For the cell lines used in this study no protocols for siRNA transfection existed in the laboratory, so this had to be established. Due to previous success with Pre-miR™ miRNA Precursor transfections, siRNA transfection optimisation was performed with the forward transfection method. For RNAi Ambion's Silencer™ Select siRNAs *(4427037, Ambion, Applied Biosystems, Life Technologies)* were used. These siRNAs are designed with Ambion's proprietary algorithm and are claimed by the manufacturer to have higher specificity and effectiveness at lower concentrations than previous generation products.

RNAi optimisation in 2102Ep and NTera-2 cells was performed with different candidate genes for each cell line. For 2102Ep cells, the SOX2 gene was used, whereas for NTera-2 cells, GAPDH was used as RNAi optimisation control.

4.4.1.1 — Gene Expression Analysis

SOX2 was the first target chosen for functional knock-down in 2102Ep cells, as its function as a crucial stem cell marker suggested its silencing could result in 2102Ep differentiation. This is detailed further on in Section 4.4.2. For this reason, the transfection optimisations were performed with the siRNA targeting this gene. For the first stage of transfection optimisation, forward transfections were performed on pre-seeded 96-well cell culture plates using the protocol described in Section 4.3.2.1. High concentration Lipofectamine™ 2000 (0.375 µl/well) were used for transfection in 2102Ep cells. The post-transfection incubation time was reduced from 48 hours, as used for Pre-miR™ miRNA Precursor transfection optimisation, to 24 hours. This was done to reduced the possibility of producing misleading SOX2 silencing results due to a potential SOX2 knock-down-induced differentiation cascade further downregulating the gene beyond the RNAi capacity of the SOX2 siRNA. This time frame was at the lower side of the 24-48 hour post-transfection period recommended by the manufacturer. Initial transfection was performed with 10 and 15 nM Silencer™ Select siRNA, which was within the 5-30 nM range recommended by the manufacturer. Gene expression analysis was performed as described in Section 4.3.2.2.

NTera-2 siRNA transfection optimisation was performed with fellow student, Salah El Baruni, initially for his research. For RNAi optimisation the gene GAPDH was used. As with Pre-miR™ miRNA Precursor Optimisation, forward transfections were performed with Medium (0.2 µl/well) and High (0.3 µl/well) volumes of Lipofectamine™ RNAiMAX on a 96-well cell culture plate. Invitrogen, the manufacturer of Lipofectamine™ transfection agents, previously tested gene knock-down in NTera-2 cells with Lipofectamine™ RNAiMAX and 30 nM Silencer™ Select siRNA against GAPDH. Transfections were performed with the protocol described in Section 4.3.3.1. Post-transfection incubation time was 48 hours. Gene expression results were produced with the method described in Section 4.3.2.2. GAPDH expression was normalised to the gene B2M, a commonly used endogenous control and previously validated for the NTera-2 cell line. Gene expression results are displayed in Figure 4.3.

SOX2 gene expression results can be found in Figure 4.2. In 2102Ep cells SOX2 silencing was considered sufficient, with a highly consistent 70% downregulation of SOX2 expression in 10 nM siSOX2 transfected cells, which was expected for the post-transfection incubation time used. As shown, a higher concentration of siSOX2 (15 nM) resulted in a decreased knock-down of SOX2, so no further optimisation with siRNA concentrations was performed.

NTera-2 gene expression results showed a significant and consistent GAPDH knock-down of 88% in cells transfected with 30 nM siGAPDH in High concentrations of Lipofectamine™ RNAiMAX. This suggests a sufficient transfection efficiency for further functional transfections.

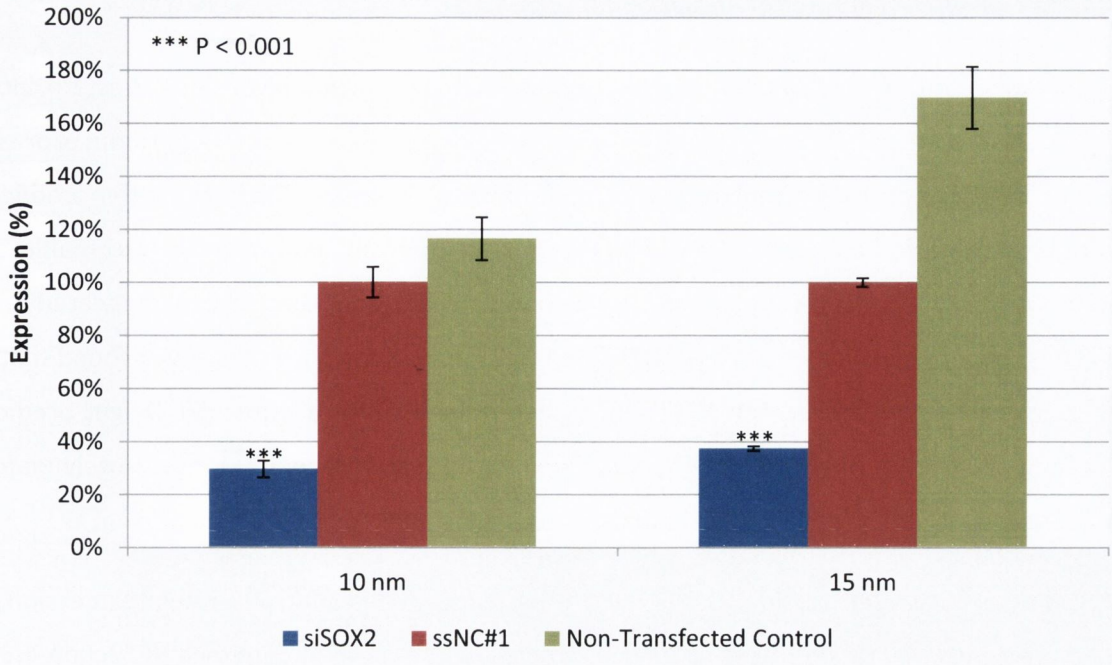


Figure 4.2—SOX2 silencing in 2102Ep cells transfected with 10 and 15 nM siSOX2. A significant downregulation of SOX2 in 10 and 15 nM siSOX2 transfected 2102Ep cells was reported compared to the ssNC#1-transfected cells ($P < 0.0001$ and $P = 0.0005$ respectively). SOX2 expression in non-transfected control cells was significantly upregulated compared to the ssNC#1-transfected cells ($P = 0.0057$). SOX2 expression in ssNC#1-transfected cells was set at 100%. All experiments were performed in biological triplicate and the values shown are the mean expression of these triplicates ± SD. A t-test was used to compare sample means.

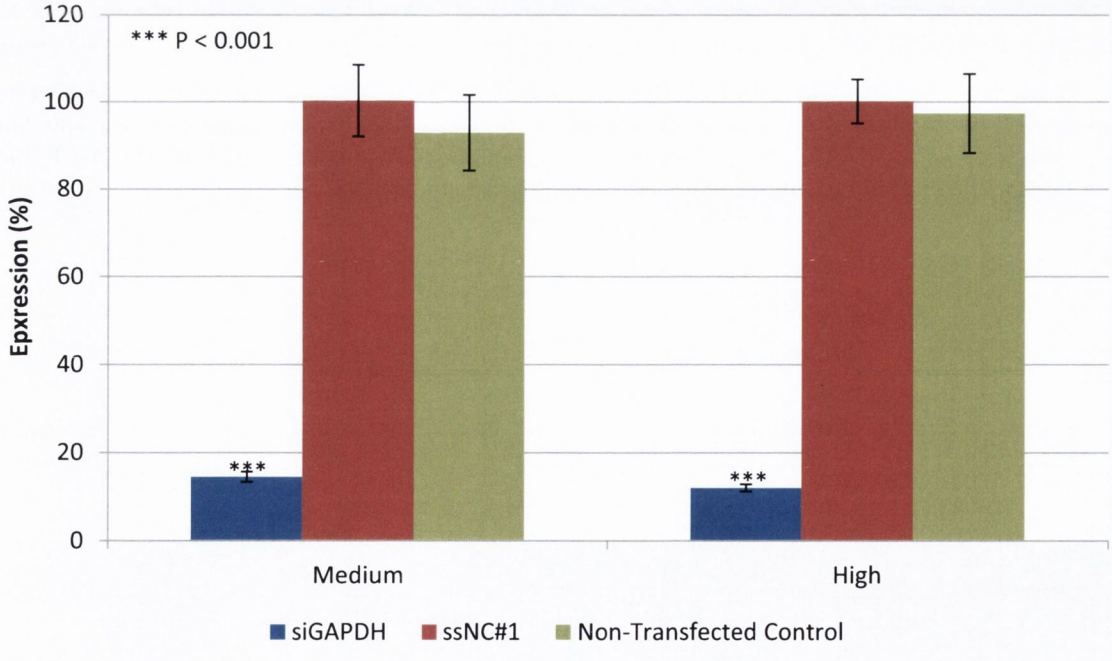


Figure 4.3—GAPDH silencing in NTera-2 cells transfected with 30 nM siGAPDH. For both High and Medium concentrations of Lipofectamine™ RNAiMAX, 30 nM siGAPDH transfection resulted in a significant downregulation of GAPDH in NTera-2 cells ($P = 0.0005$ and $P < 0.0001$ respectively). The expression in the ssNC#1 samples was set at 100%. All experiments were performed in biological triplicate and the values shown are the mean expression of these triplicates ± SD. A t-test was used to compare sample means.

4.4.1.2 — Protein Expression Analysis

SOX2 expression in 2102Ep cells was also assessed at the protein level using Western blot as described in Section 4.3.2.3. Figure 4.4 demonstrates a clear silencing of SOX2 protein expression in all three siSOX2 transfected replicates. SOX2 protein expression does not differ among the ssNC#1-transfected samples and the non-transfected controls, indicating no discernable side-effects on SOX2 expression due to the transfection process. A unexpected band appeared on the blot at a lower molecular weight than SOX2. As SOX2 has no known isomers, it was found that this band most closely matched SOX3 protein, as determined from a molecular weight prediction (P41225, UniProtKB/Swiss-Prot).[438] The SOX3 amino acid sequence was the most closely matched sequence to SOX2 among all known human genes (blastp, BLAST®, NCBI).[439]

In addition to gene expression, GAPDH and B2M (endogenous control) protein expression was measured with the Western blot technique, employing the method described in Section 4.3.2.3. This was performed with cells cultured and transfected on 6-well plates with a High concentration of Lipofectamine™ RNAiMAX (8.46 µl/well).

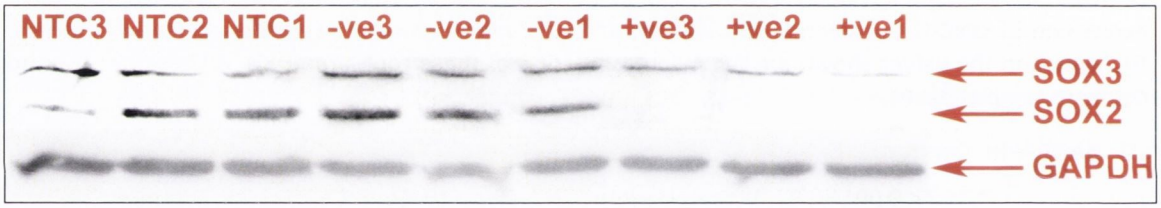


Figure 4.4—Western blot of SOX2 silencing in 2102Ep cells. Cells transfected with siSOX2 (+ve) showed clear silencing of SOX2 when compared to the ssNC#1-transfected cells (-ve) and non-transfected controls (NTC). SOX2 protein was silenced in cells transfected with siSOX2 (+ve1, +ve2 and +ve3). The additional band above SOX2 was predicted to be SOX3.

4.4.1.3 — Cytotoxic Effects of siRNA Transfection

Cell viability experiments were performed with ssNC#1 to determine the impact of the transfection process on cell proliferation.

For 2102Ep cells this was performed with 10 nM ssNC#1-transfected at High Lipofectamine™ 2000 concentrations. Low and Medium Lipofectamine™ 2000 transfected cells and untreated cells were included as references. For NTera-2 cells this was performed by transfecting 30 nM ssNC#1 with Low, Medium and High concentrations of Lipofectamine™ RNAiMAX. For both cell lines cell viability was assessed 48 hours after transfection with the method described in Section 2.1.8.

Cell viability is displayed in Figure 4.5. This shows that the transfection process can impact cell viability, but this can depend on the concentration of transfection agent used. In 2102Ep cells, only ssNC#1-transfected cells with a Medium concentration (0.25 µl/well) of Lipofectamine™ 2000 appear to have significantly decreased viability (22% reduction). NTera-2 cells transfected with all Lipofectamine™ RNAiMAX concentrations showed a significant reduction in cell viability compared to the untreated cells. These reductions are between 14%-24%. Cells transfected with High Lipofectamine™ RNAiMAX concentrations seemingly experienced the least amount of cytotoxicity (14%). Experimental siRNA transfections were to be conducted at this concentration and 14% reduction of cell viability was considered acceptable.

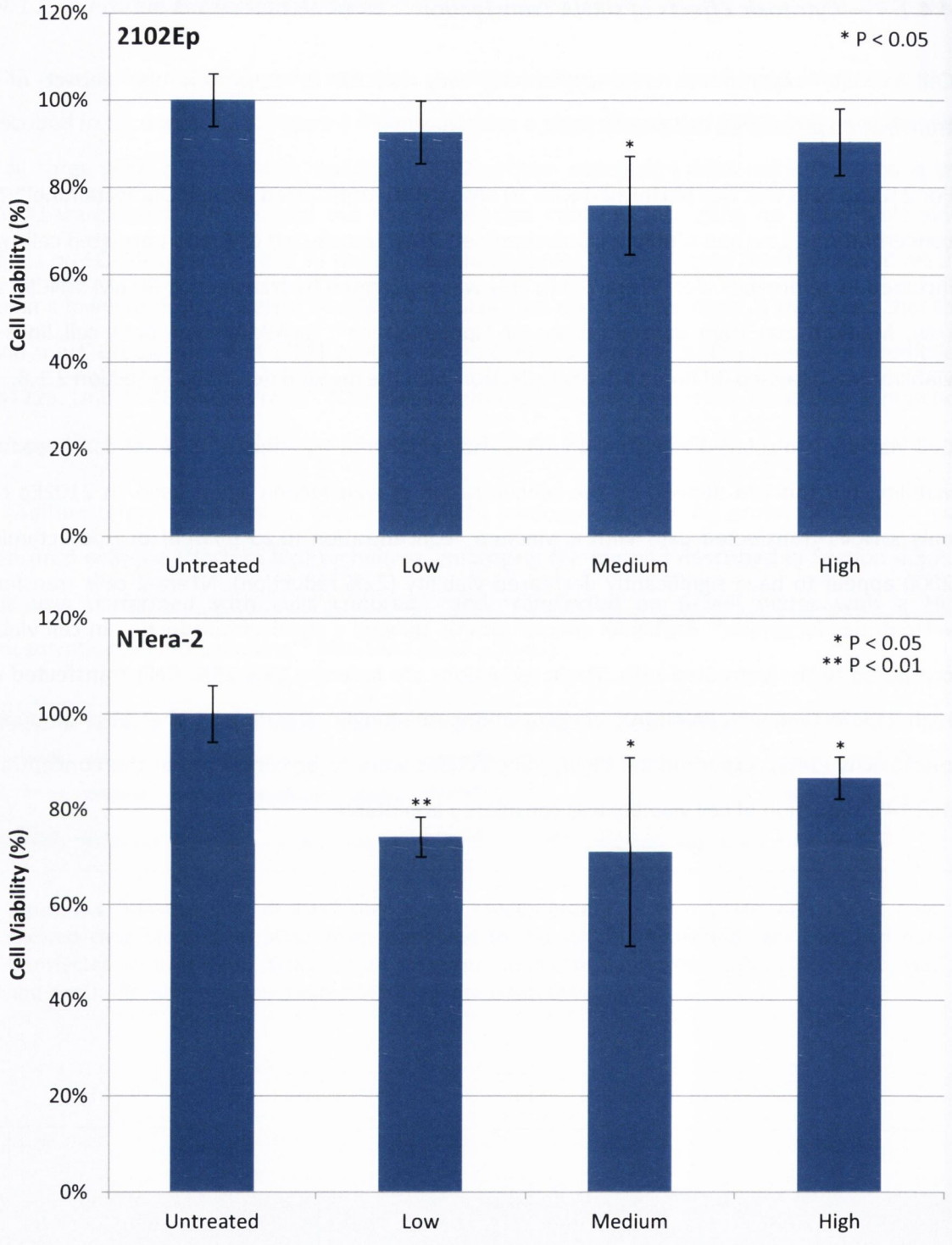


Figure 4.5—Cell viability of 2102Ep and NTera-2 cells transfected with ssNC#1 at three different concentrations of Lipofectamine™ 2000 and Lipofectamine™ RNAiMAX respectively. In 2102Ep cells, ssNC#1-transfected with Medium concentration of transfection agent significantly reduced cell viability ($P = 0.0316$). Cell viability was significantly lower in all three transfected NTera-2 samples (Low: $P = 0.0026$, Medium: $P = 0.0304$, High: $P = 0.0295$). The cell viability of the untreated controls was set at 100%. All experiments were performed in biological triplicate and the values shown are the mean cell viability of these triplicates ± SD. A t-test was used to compare sample means.

4.4.2 — Functional Silencing of SOX2 in 2102Ep Cells

The first candidate for silencing in 2102Ep cells was SOX2. 2102Ep cells are resistant to ATRA-induced differentiation. One of the aims of this study was to find novel methods to induce differentiation in 2102Ep cells. For first step to investigate this, SOX2, as a crucial stem cell marker, was selected for RNAi-induced silencing in 2102Ep cells with the aim of inducing spontaneous differentiation. For this, SOX2 was silenced with 10 nM siSOX2 in 2102Ep cells on 6-well plates using the protocol described in Section 4.3.3.1. After transfection the cells were incubated for three days.

4.4.2.1 — Phenotypic Analysis

During transfection changes in cell morphology were registered by photographing the cells under a light microscope. The images in Figure 4.6 were taken at three days post-transfection. The siSOX2 transfected 2102Ep cells showed a very distinct morphology from the cells transfected with ssNC#1. The cells had a flatter and larger appearance. The cells were also more homogenously spread around the plate, which differed from their normal appearance as densely packed colonies. The cells were quite similar in appearance to differentiating NTera-2 cells.

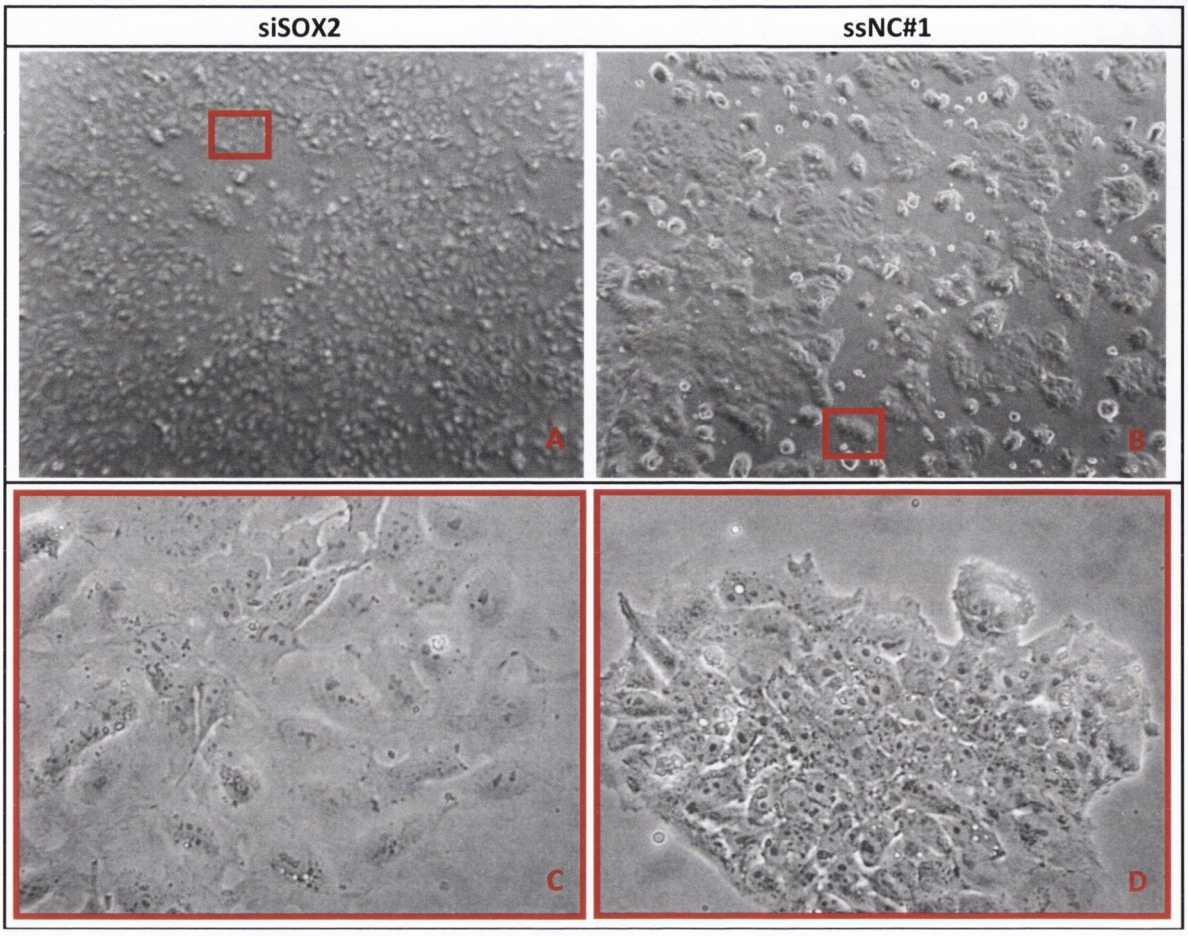


Figure 4.6—Microscopic images of 2102Ep cells transfected with siSOX2 and ssNC#1. A dramatic morphological change can be seen for cells transfected with siSOX2 (A and C) compared to those transfected with ssNC#1 (B and D). The cells became flatter, enlarged and less dense. Images taken at 50X (A and B) and 320X (C and D) The red border indicates similar areas of similar size.

4.4.2.2 — Gene Expression Analysis

Gene expression of pluripotency markers, SOX2, OCT4 and NANOG, was measured and is shown in Figure 4.7. SOX2 silencing resulted in a significant suppression of all three markers. As expected, SOX2 saw the highest downregulation, with only 8% remaining expression compared to the cells transfected with ssNC#1. NANOG and OCT expression was at 35% and 51% respectively. Unexpectedly, when compared to the ssNC#1-transfected cells and the vehicle control, NANOG expression was significantly lower in the non-transfected control cells ($P = 0.0418$), although this is not as substantial as the downregulation measured in the siSOX2 transfected cells. In addition, when these two latter controls are compared, NANOG expression in siSOX2 transfected cells is significantly downregulated ($P = 0.0193$).

Differentiation marker expression showed in Figure 4.8 indicated a significant 4.2-fold upregulation of NCAM1 and a significant 109.3-fold upregulation of GATA6 expression in siSOX2 transfected 2102Ep cells. Conversely, ENO3 is significantly downregulated by 40%. Unusually, NCAM1 and ENO3 expression is significantly downregulated in the non-transfected controls

compared to the ssNC#1-transfected cells (P = 0.0260 and P = 0.0259). Together, all results and observations made suggest strongly that spontaneous differentiation was occurring in 2102Ep cells. GATA6, as a trophectodermal and endodermal marker, is highly upregulated, suggesting the cells were differentiating towards these dermal phenotypes. This agrees with previous research which shows 2102Ep differentiation towards trophectodermal cell lineages upon the silencing of OCT4.[297]

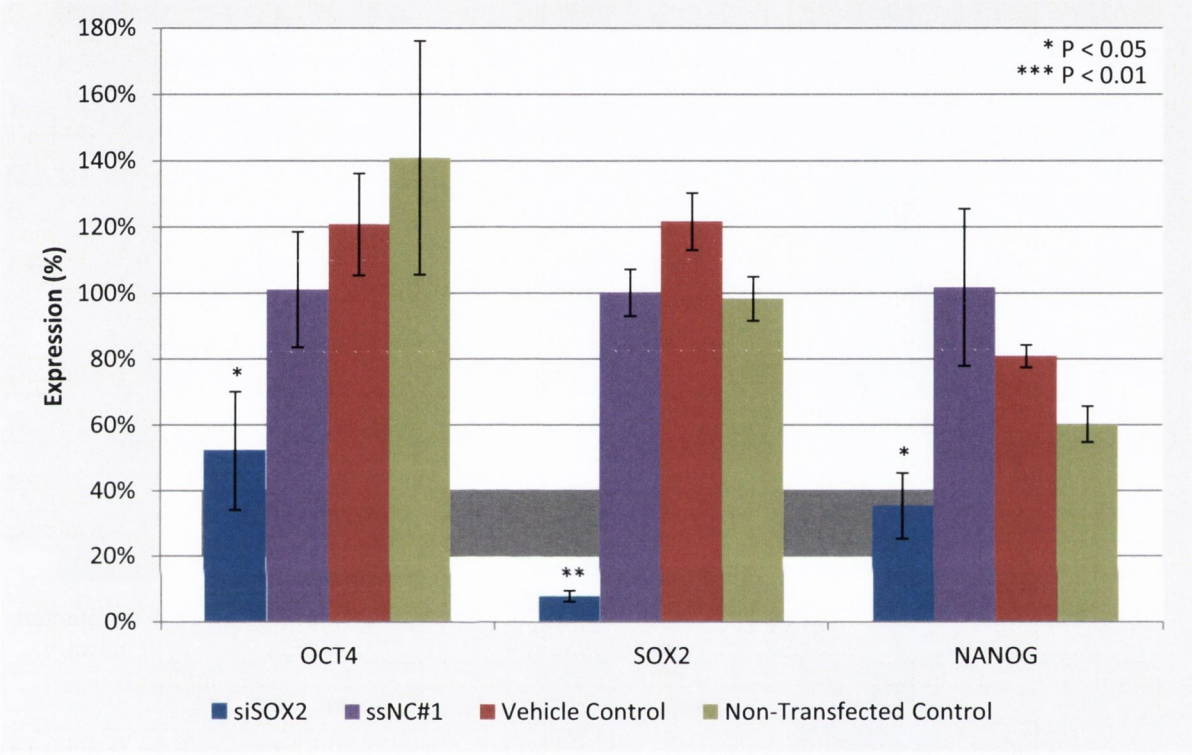


Figure 4.7—Stemness marker expression in 2102Ep cells transfected with siSOX2. OCT4, SOX2 and NANOG all significantly downregulated in siSOX2 transfected cells compared to ssNC#1-transfected cells (P = 0.0173, P < 0.0001 and P = 0.0111 respectively). The expression in the ssNC#1 samples was set at 100%. All experiments were performed in biological triplicate and the values shown is mean relative expression of these triplicates ± SD. A t-test was used to compare sample means.

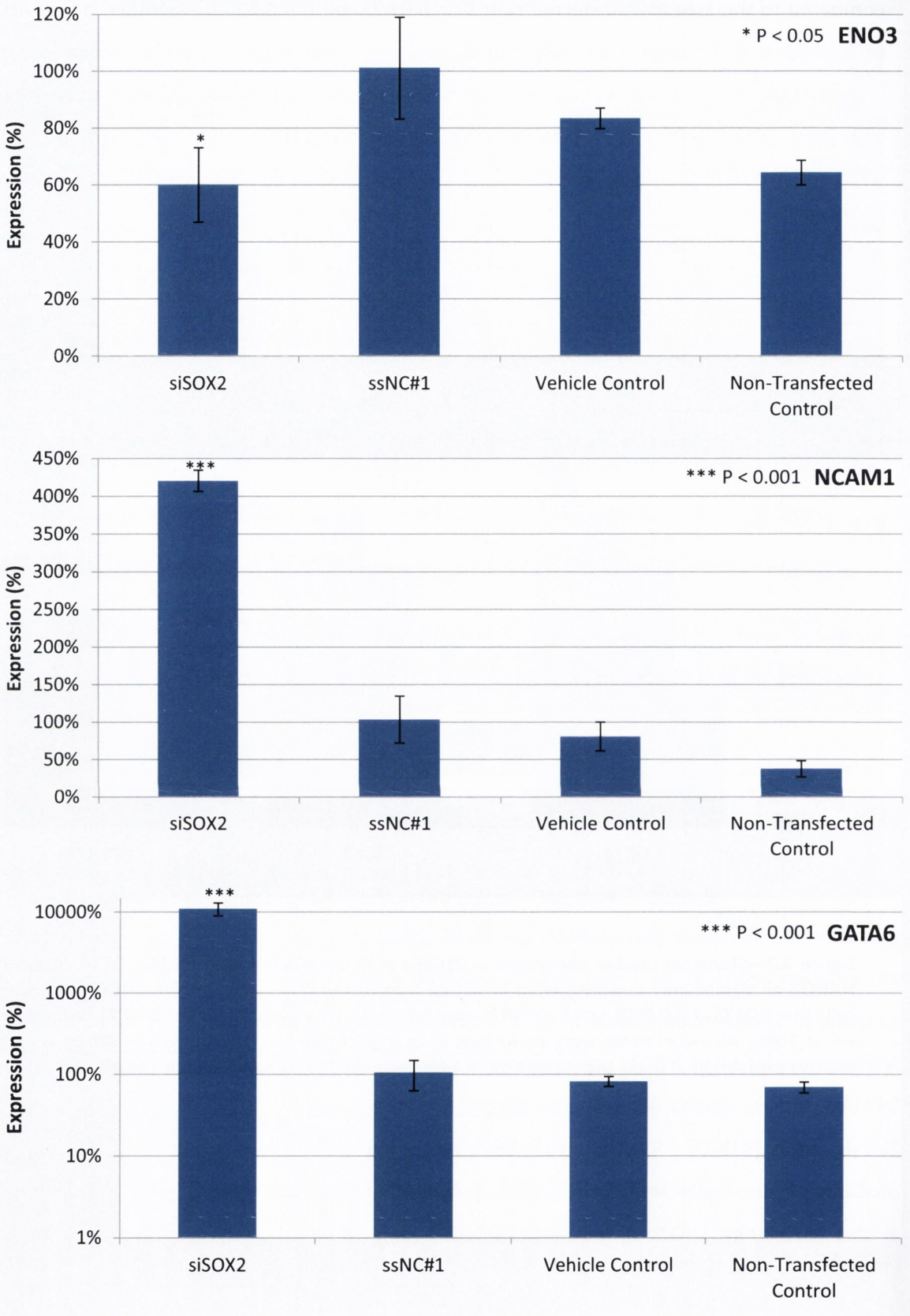


Figure 4.8—Differentiation Marker Expression in siSOX2 transfected 2102Ep cells. ENO3 is significantly downregulated in siSOX2 transfected cells compared to ssNC#1-transfected cells (P = 0.0326). Conversely, NCAM1 and GATA6 are significantly upregulated (P = 0.0010 and P = 0.0007 respectively). The expression in the ssNC#1 samples was set at 100%. All experiments were performed in biological triplicate and the values shown are the mean expression of these triplicates ± SD. A t-test was used to compare sample means.

4.4.2.3 — Assessing Differentiation with an Alkaline Phosphatase Assay

Undifferentiated ESCs and ECCs highly express cell surface APs, which are reduced upon differentiation of these cells.[284] Both 2102Ep and NTera-2 cells highly express cell surface APs and in NTera-2 cells, the expression of these is reduced upon retinoic acid-induced differentiation.[284,314] The enzymatic activity of AP can be measured with enzymatic assays which can stain cells with high cell surface AP expression, as described in Section 2.1.9. In addition to measuring mRNA expression of pluripotency and differentiation markers, as described in the previous Sections, SOX2-silenced 2102Ep cells were also subjected to an AP staining assay. This was performed for two time points: four and 7 days after the transfection of SOX2 siRNA. Transfections were performed on 24-well plates as described in Section 2.2.2.5 and the AP staining assay was performed by Section 2.1.9 after initial optimisation. AP staining of 2102Ep cells for four and 7 days is shown in Figure 4.9. After four and 7 days, AP expression was notably reduced in cells transfected with siSOX2, although it was not completely gone. An aberrant differentiation program prevents 2102Ep cells to completely differentiate, which could prevent the complete loss of AP expression. Alternatively, the 4-7 day period could be too short to allow for substantial cell surface AP loss.[41,296] As a comparison, NTera-2 cells exposed to 10^{-2} mM ATRA showed an apparent complete loss of cell surface AP after four days as detected with the same AP assay solution. These results are shown in Figure 4.10.

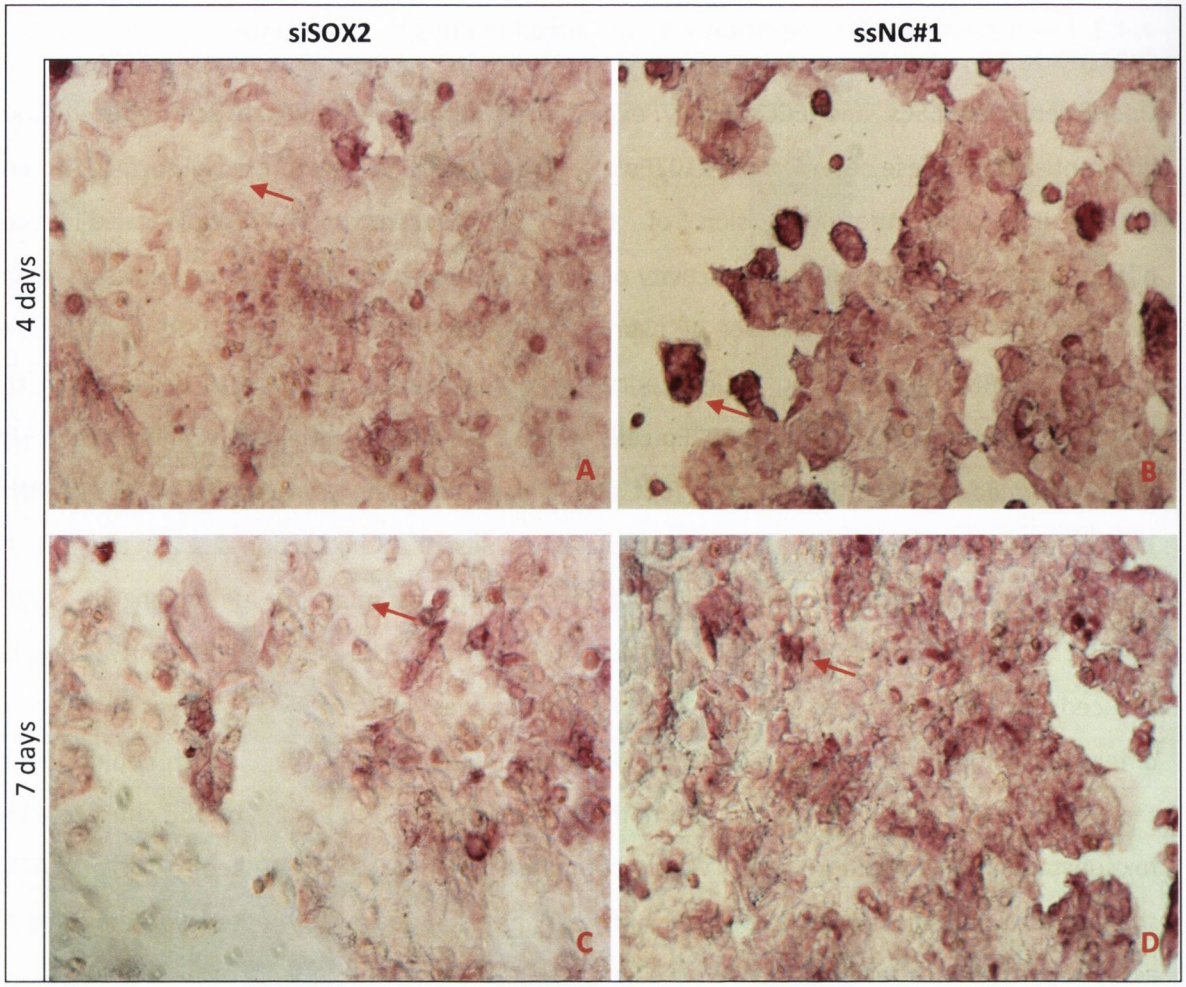


Figure 4.9—Staining of AP on 2102Ep cells transfected with siSOX2 and ssNC#1 and incubated for four and 7 days. AP expression was visibly reduced by SOX2 silencing, as can be seen by a reduced red staining in cells transfected with siSOX2 (A and C). Comparatively, regions of intense red staining was seen in particular cell colonies in samples transfected with ssNC#1 (B and D). Red arrows indicate examples of observations described. Images were taken at 100X magnification.

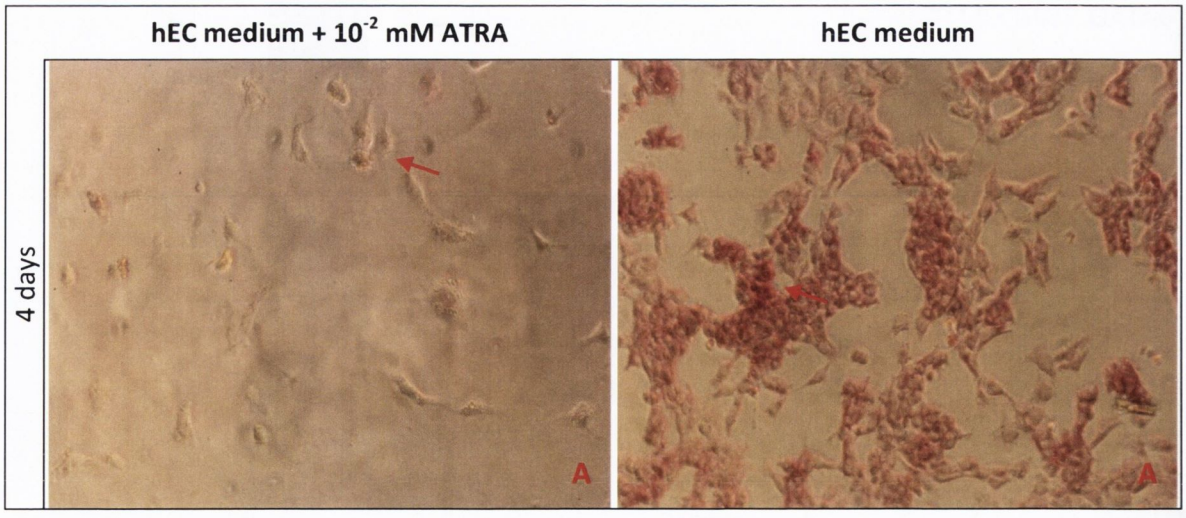


Figure 4.10—Staining of AP on NTera-2 cells exposed to 10^{-2} mM ATRA for four days. Cells exposed to ATRA lost AP expression as marked by a complete loss of red staining (A). Samples left in growth medium retained an intense red staining indicating substantial expression of AP in undifferentiated cells (B). Red arrows indicate examples of observations described. Images were taken at 100X magnification.

4.4.3 — Functional Silencing of SOX2 in NTera-2 Cells

Functional silencing of SOX2 in NTera-2 cells was performed after the optimisation of GAPDH knock-down (see Section 4.3.2). In triplicate, 30 nM SOX2 siRNA was transfected into the cells on 6-well plates with the method described in Section 4.3.3.1. For ssNC#1 samples, 30 nM ssNC#1 was similarly transfected and vehicle and non-transfected controls were included.

4.4.3.1 — Phenotypic Analysis

During transfection changes in cell morphology were registered by photographing the cells under a light microscope. The images in the first row Figure 4.11 were taken at three days post-transfection. This showed little morphological change in the siSOX2-transfected cells compared to the negative controls. Possibly, the morphological effects were hidden by the high density of the cells. The experiment was repeated and the post-transfection incubation time was extended to 6 days, with a cell split after three days. The difference in cell morphology was greatest at four days, when cell confluency was still relatively low. The images in the first and second row of Figure 4.11 clearly showed a dispersion of the cells and a flattened and enlarged morphology with multiple branching projections, compared to the small, dense and uniform morphology of the negative controls.

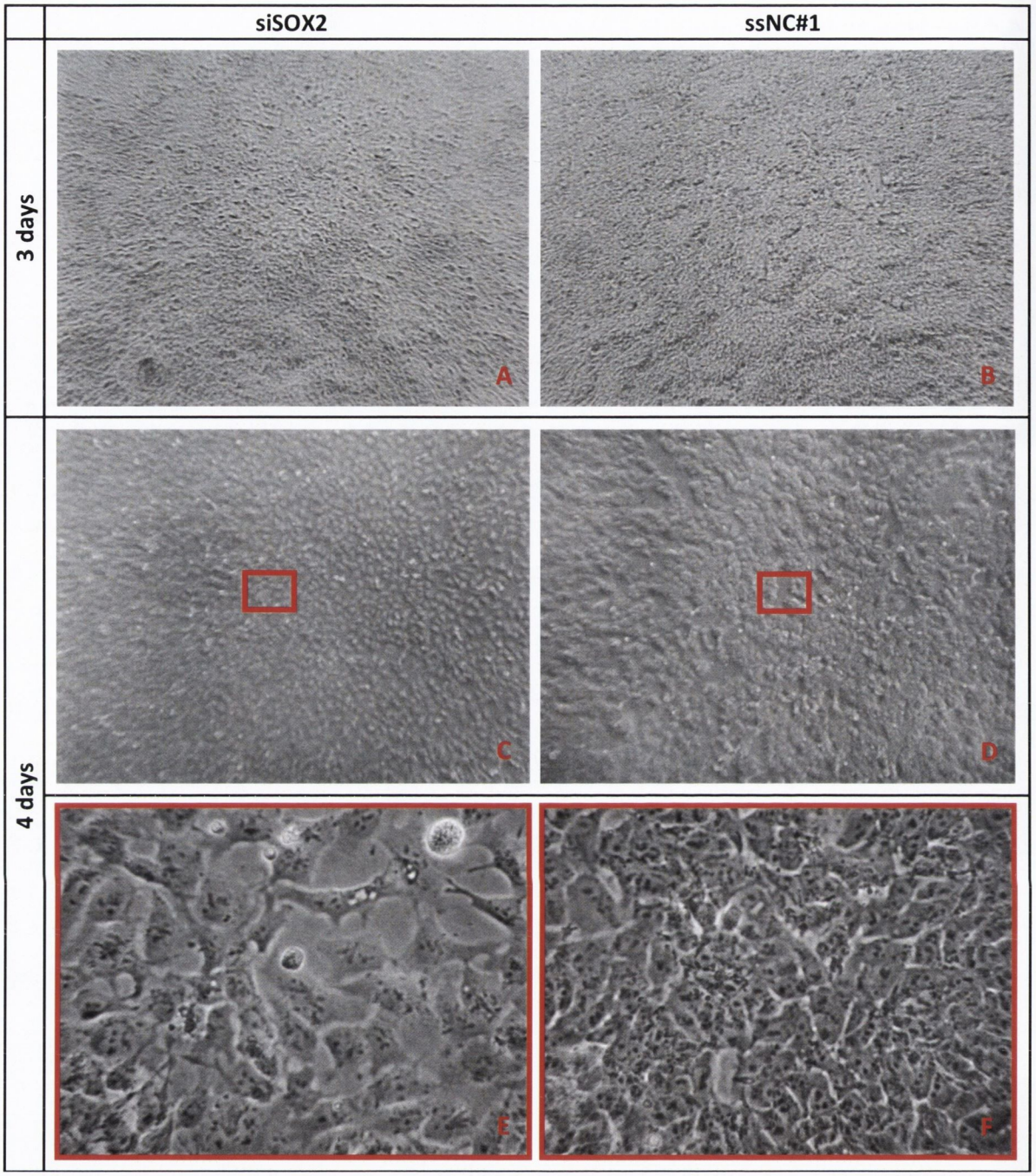


Figure 4.11—Microscopic images of NTera-2 cells transfected with siSOX2 and ssNC#1. Differences in cell morphology were hard to identify after three days due to high cell density (A and B). After a cell passage, a flatter and enlarge morphology was clearly seen at four dayas post-transfection (C and D). This is more visible at higher magnification (E and F). Images taken at 50X (A, B, C and D) and 320X (E and F). The red border indicates similar areas of similar size.

4.4.3.2 — Gene Expression Analysis

The gene expression of the pluripotency markers, SOX2, OCT4 and NANOG, was analysed, as by the method in Section 4.3.2.2. The expression of these were measured at three and 6 days after transfection of siSOX2. The results of these are listed in Figure 4.12. These show that SOX2 is significantly downregulated by 51% by siSOX2. NANOG was significantly downregulated by 73%, but OCT4 expression was not reduced significantly. These results were unexpected, as 51% SOX2 knock-down was considered quite inefficient. An additional optimisation step with 5, 10 and 15 nM siRNA concentrations did not improve SOX2 knock-down, as is shown in Figure 4.13. An siRNA concentration of more than 30 nM resulted in an unacceptable degree of cytotoxicity (data not shown). OCT4 expression did not respond to SOX2 downregulation after three days, but because NANOG expression was substantially reduced, it was thought that a degree of differentiation might be occurring. To test whether siSOX2 transfection could indeed result in the significant downregulation of pluripotency factors, the post-transfection incubation time was extended to 6 days. The results from this, shown in Figure 4.14, clearly showed a significant and substantial downregulation of all three pluripotency markers, particularly SOX2 and NANOG, which were almost completely silenced.

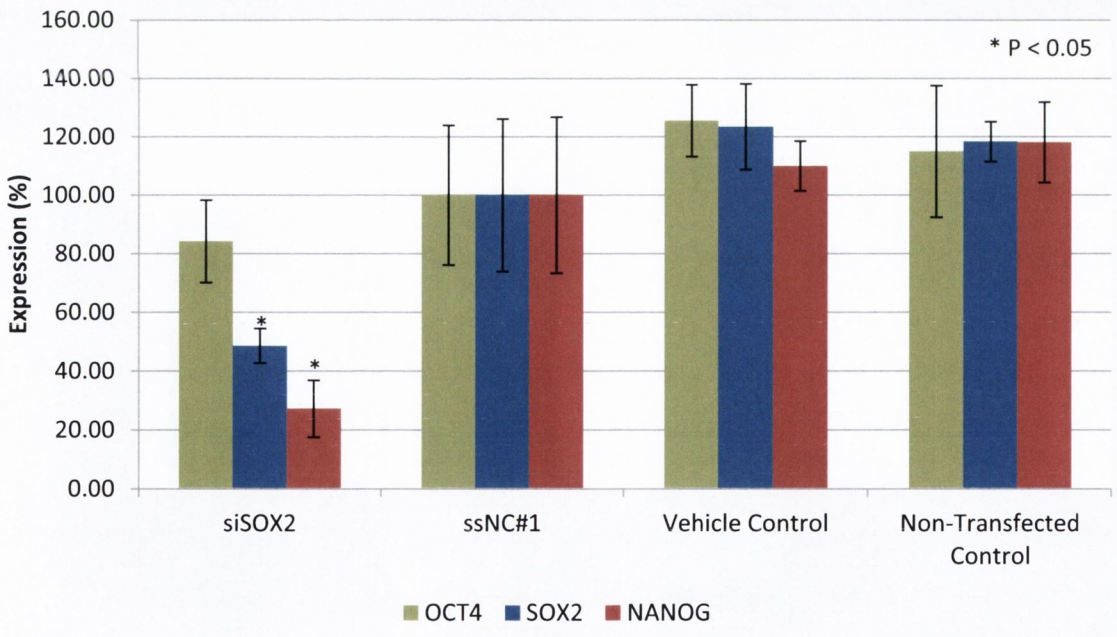


Figure 4.12—Pluripotency marker expression in NTera-2 cells three days after transfection with 30 nM siSOX2. SOX2 and NANOG expression were significantly lower in siSOX2 transfected cells compared to the ssNC#1-transfected cells (P = 0.0252 and P = 0.0102 respectively). The expression in the ssNC#1 samples was set at 100%. All experiments were performed in biological triplicate and the values shown are the mean expression of these triplicates ± SD. A t-test was used to compare sample means.

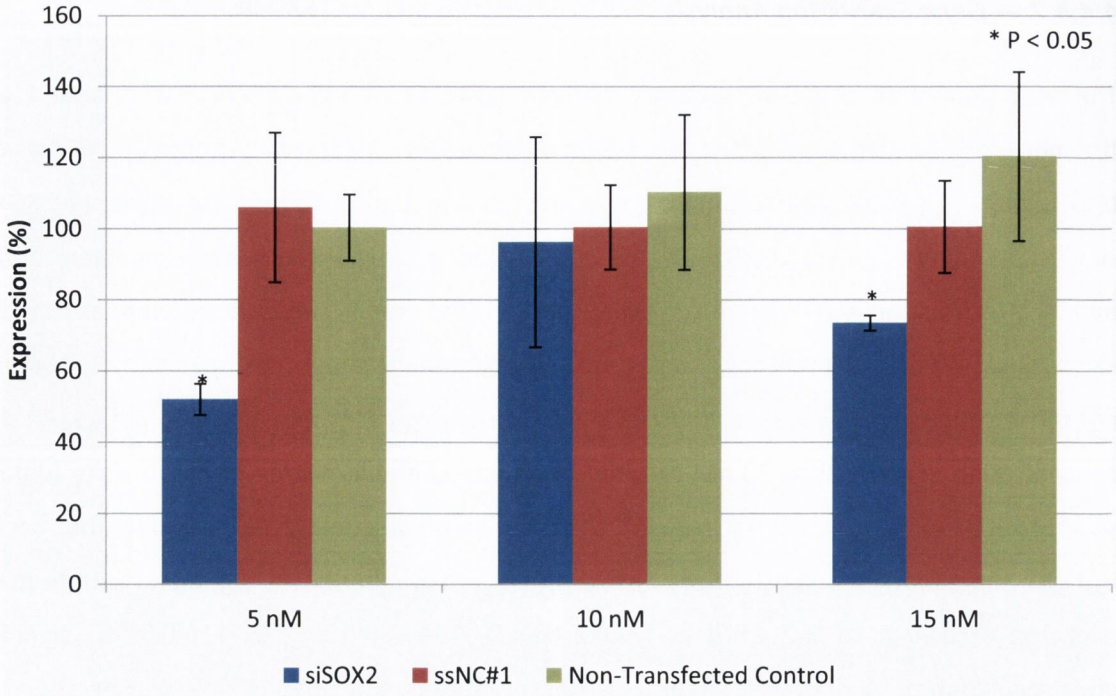


Figure 4.13—SOX2 expression in NTera-2 cells transfected with 5, 10 and 15 nM siSOX2. SOX2 expression was significantly lower in cells transfected with 5 and 15 nM siSOX2 compared to the ssNC#1-transfected cells (P = 0.0121 and P = 0.0234 respectively). The expression in the ssNC#1 samples was set at 100%. All experiments were performed in biological triplicate and the values shown are the mean expression of these triplicates ± SD. A t-test was used to compare sample means.

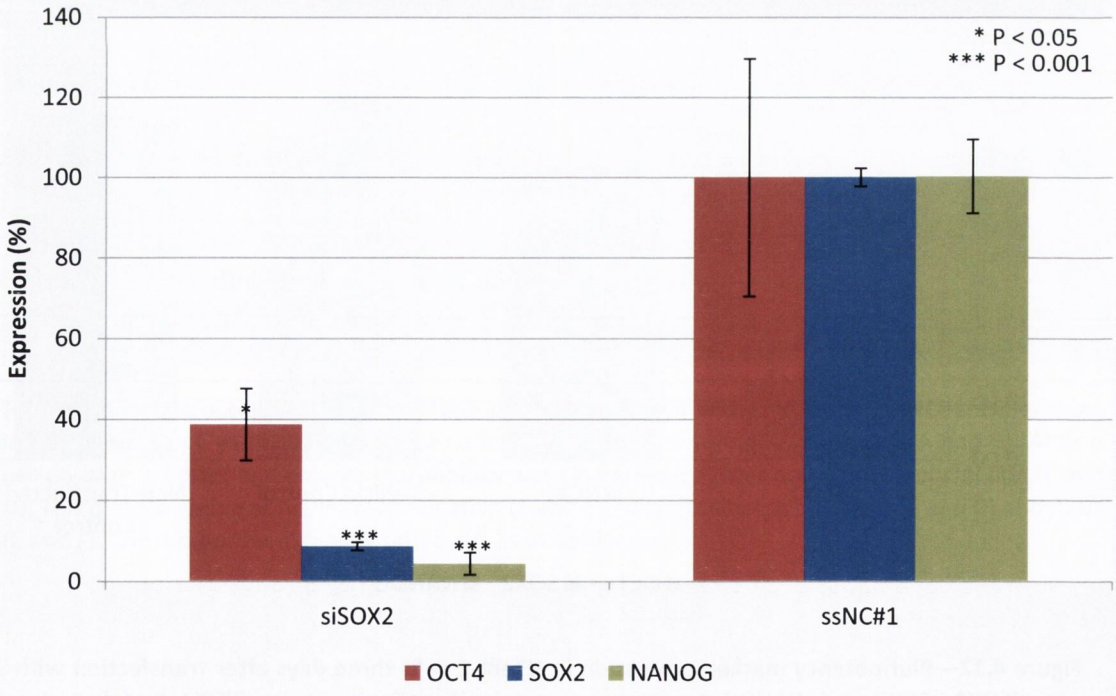


Figure 4.14—Pluripotency marker expression in NTera-2 cells at 6 days after transfection with 30 nM siSOX2. SOX2, OCT4 and NANOG expression were significantly lower in siSOX2 transfected cells compared to the ssNC#1-transfected cells (P < 0.0001, P = 0.0446 and P = 0.0008 respectively). The expression in the ssNC#1 samples was set at 100%. All experiments were performed in biological triplicate and the values shown are the mean expression of these triplicates ± SD. A t-test was used to compare sample means.

4.4.3.3 — Assessing Differentiation with an Alkaline Phosphatase Assay

Similarly to 2102Ep cells, cell surface AP expression was measured on NTera-2 cells using the protocol from Section 2.1.9. The results, shown in Figure 4.15, indicate a strong reduction of cell surface AP expression. When compared to AP reduction in NTera-2 cells exposed to 10^{-2} mM ATRA, as shown in Figure 4.10, the reduction is less. However, the siSOX2-transfected culture is denser, suggesting that this might participate comparatively in more intense AP staining.

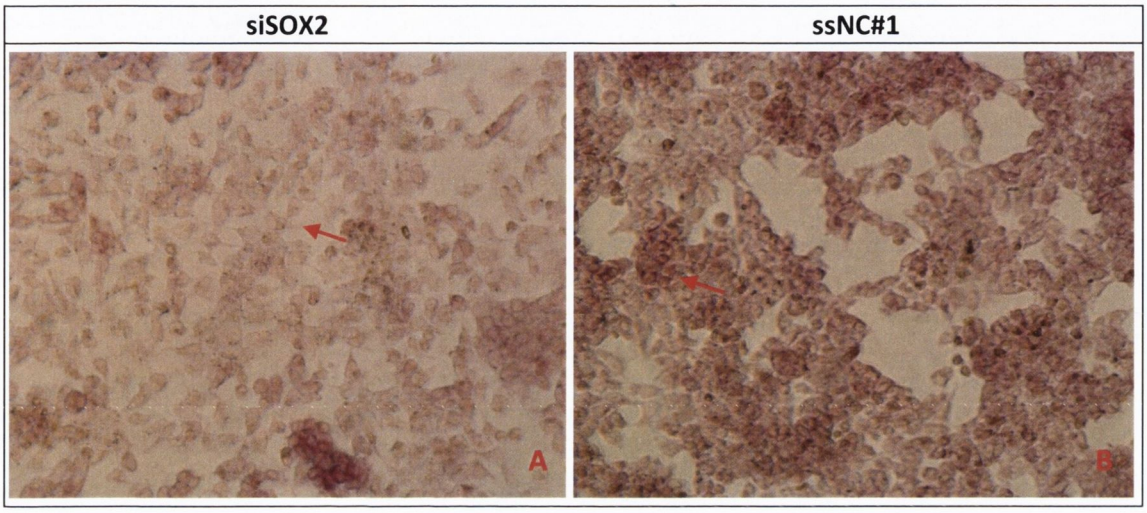


Figure 4.15—Staining of AP on NTera-2 cells transfected with siSOX2 and ssNC#1 and incubated for four days. AP expression was visibly reduced by SOX2 silencing, as can be seen by a reduced red staining in cells transfected with siSOX2 (A), compared to cells transfected with ssNC#1, which showed intense staining (B). Red arrows indicate examples of observations described. Images were taken at 100X magnification.

4.4.4 — Functional Silencing of Five Genes in 2102Ep

The aim of this chapter was to find novel genes which play a role in the differentiation of embryonal carcinoma cells. The goal was to find targets and pathways which could potentially be unique to the differentiation of these cells and which would be distinguished from the differentiation of hESC, with the intention to find unique genes which potentially promote the aberrant behaviour of CSCs.

Previous studies in the laboratory produced whole-genome expression data after three days of the ATRA-induced differentiation of NTera-2 cells. In NTera-2 cells, 454 genes were significantly downregulated and 312 genes were upregulated by a fold change of ≥2. The list of top downregulated genes were ranked by decreasing fold-change and the list was filtered by a set of criteria:

1. Genes which were common hESC markers according to Cai et al. were excluded.[335]
2. Genes which were downregulated in 1-6 and BG01 hESC.[440]
3. Genes with no known other previously validated function in hESC or ECC stemness.
4. Silencer Select™ and primary antibodies were available for these genes.

The initial experiment design targeted approximately 100 candidate genes for silencing with siRNA on 96-well plates on a high-throughput scale. An AP assay stain, as used in Section 4.4.2.3, was intended to be used to assess the induction of cell differentiation. It was prohibitively difficult to detect a sufficient difference in AP colouration between differentiated and undifferentiated 2102Ep and NTera-2 cells after three days of post-transfection incubation. A sufficient contrast in AP staining was detected after four days of post-transfection incubation, which required splitting the cells at three days. This was considered exceedingly difficult to integrate in a high-throughput protocol, while assuring consistency. Therefore, a list of five genes was generated to be targeted for RNAi-induced knock-down in both 2102Ep and NTera-2 cells. These genes and their known properties and related research can be found in Table 4.9.

Transfections were performed on 6-well plates as described in Section 4.3.3.1. Cells were incubated for three days post-transfection during which images were taken with a light microscope.

Table 4.9—Previous established functions of the five genes selected for silencing in 2102Ep and NTera-2 cells. In addition to functional information, the fold downregulation of these genes in NTera-2 cells exposed to 10^{-2} ATRA for three days is denoted.

VSNL1 (103.5-fold)
· Neuronal calcium sensor protein. Part of the cAMP-dependent signalling pathway. · An oncogene which is upregulated in high invasive neuroblastoma.[441] · Moderately well expressed in 2102Ep cells and overexpressed compared to NSCs.[280] · Highly expressed in embryonal carcinoma tissue compared to differentiated tissue.[442] · Downregulated in oesophageal tumours.[443] · This gene is very stable in three types of differentiating hESCs.[444] · Downregulation induces cAMP-dependent differentiation in rat glioma cells.[445] · Enhances cAMP levels by PKA induction and inhibits squamous carcinoma migration.[446] · Also downregulated (1.6-fold) in 2102Ep cells upon RA-induction.
DPPA3 (44.8-fold) **OS**
· Probably some sort of transcription factor/repressor · Germ cell marker. · This gene may play a role in transcriptional repression, cell division, and maintenance of cell pluripotency. · Highly expressed in 2102Ep cells and extremely overexpressed compared to NSCs.[280] · Involved in embryonal development · Mouse ESCs differentiate into primordial germ cells when this gene is upregulated.[447] · Regulated by OCT4 which targets the genes NANOG locus in mouse oocytes.[448] · Expressed seminoma and breast carcinoma.[449] · When HUMSCs cells differentiate into germ cells the expression of this gene increases and OCT4 decreases.[450]
AK4 (AK3L1) (26.7-fold) **N** **OSN**
· Probably an adenylate kinase located to the mitochondrial matrix. · Moderately well expressed in 2102Ep cells and downregulated compared to NSCs.[280] · A vulnerable kinase in multiple melanoma.[451] · Overexpressed in late stage lung cancer and it's downregulation inhibits tumour invasion.[452] · Aberrantly expressed in mouse ESC-cloned blastocysts.[453] · Also downregulated in 2102Ep cells upon RA-induction (1.6-fold).
TMEM46 (SHISA2) (26.0-fold)
· Transmembrane protein. · Very low and similarly expressed in 2102Ep compared to NSCs (LOC387914).[280] · Modulator of FGF and Wnt signalling in chick embryo and important in development.[454] · Suppresses the maturation process of FGF and Wnt receptors and contributes to cell growth and invasion.[455] · Generally little known about this gene.
RBM47 (FLJ20273) (15.4-fold) **OSN**
· Moderately low expressed in 2102Ep cells and downregulated compared to NSCs.[280] · Almost nothing known about this gene. · Might play a role in breast cancer.[456]
Down in NCCIT ECCs upon knockdown of OCT4 (O), SOX2 (S) or NANOG (N) according to Greber et al.[378]
Genes highly expressed in Embryonal Carcinoma tissue according to Skotheim et al.[442]
Genes also downregulated in 2102Ep cells.
Genes with promoter binding regions for OCT4 O), SOX2 (S) and NANOG (N) according to Boyer et al.[79]

4.4.4.1 — Phenotypic Analysis

Images were taken at three days post-transfection with a light microscope. These images are shown in Figure 4.16. 2102Ep cells transfected with siSOX2 showed a clear morphological change, from a small, densely packed and defined phenotype, to more flattened larger and homogenously distributed cells. None of the other functionally transfected cells showed a similar morphology, however cell densities did differ suggesting differences in cell proliferation. A reduction in proliferation was observed in cells transfected with siDPPA3 and an increase in density was observed in cells transfected with VSNL1. Cell density is somewhat difficult to see in the images due to heterogeneous distribution of the cells in the well.

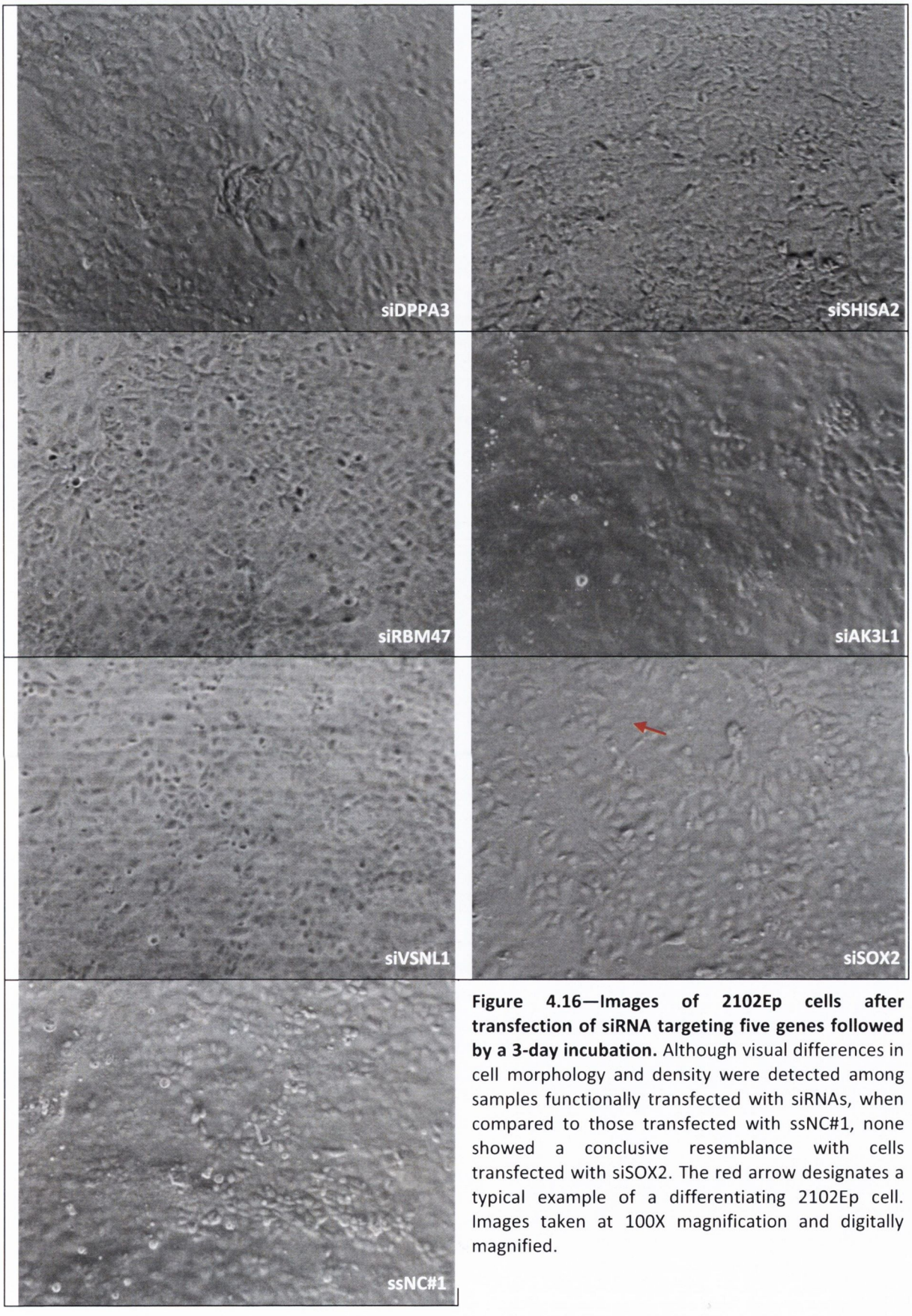
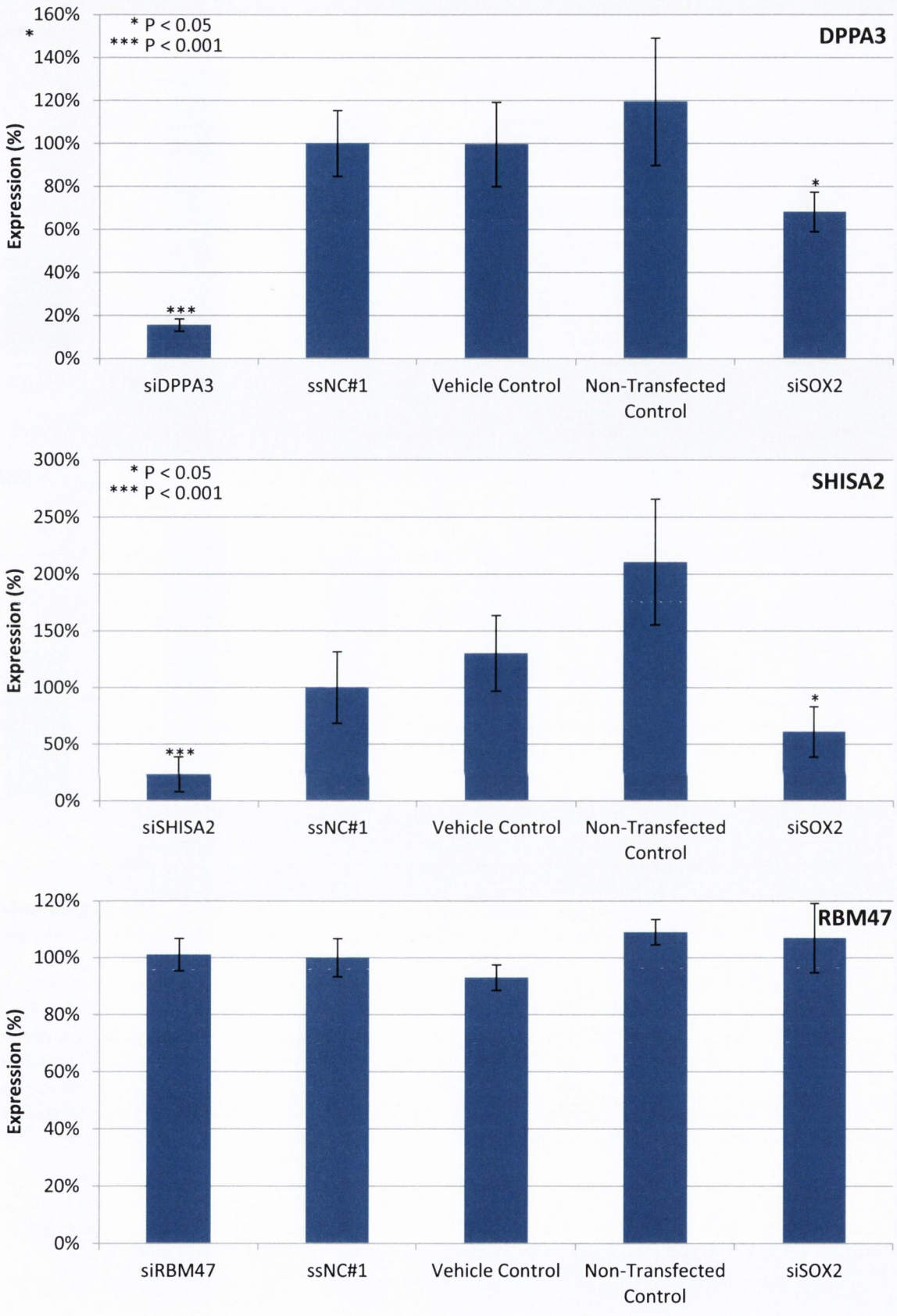


Figure 4.16—Images of 2102Ep cells after transfection of siRNA targeting five genes followed by a 3-day incubation. Although visual differences in cell morphology and density were detected among samples functionally transfected with siRNAs, when compared to those transfected with ssNC#1, none showed a conclusive resemblance with cells transfected with siSOX2. The red arrow designates a typical example of a differentiating 2102Ep cell. Images taken at 100X magnification and digitally magnified.

4.4.4.2 — Gene Expression Analysis

Analysis of gene expression in 2102Ep cells was performed with the method in Section 4.3.3.2. The expression of the five genes selected for silencing is displayed in Figure 4.17.

When compared to the ssNC#1-transfected 2102Ep cells, significant target gene knock-down was achieved with 77-85% efficiency in all cells except in those transfected with siRBM47. RBM47 has very recently been validated and no functional studies have been performed with it. The siRNA targeting this gene was not validated and no knock-down was reported. DPPA3, SHISA2, AK3L1 and VSNL1 were significantly downregulated in siSOX2 transfected cells by 32-4-% ($P = 0.0366$, $P = 0.0393$, $P = 0.0003$ and $P = 0.0053$ respectively). Additionally, SHISA2 in non-transfected controls was 2-fold upregulated compared to ssNC#1-transfected cells ($P = 0.0393$). This suggests a negative effect on the expression of this gene due to the transfection process. The expression of the five genes in the non-transfected controls ranked relative to each other can be found in Figure 4.18.


DPPA3
* P < 0.05
*** P < 0.001
Expression (%)
160%
140%
120%
100%
80%
60%
40%
20%
0%
*
***
*
siDPPA3
ssNC#1
Vehicle Control
Non-Transfected Control
siSOX2
SHISA2
* P < 0.05
*** P < 0.001
Expression (%)
300%
250%
200%
150%
100%
50%
0%
***
*
siSHISA2
ssNC#1
Vehicle Control
Non-Transfected Control
siSOX2
RBM47
Expression (%)
120%
100%
80%
60%
40%
20%
0%
siRBM47
ssNC#1
Vehicle Control
Non-Transfected Control
siSOX2

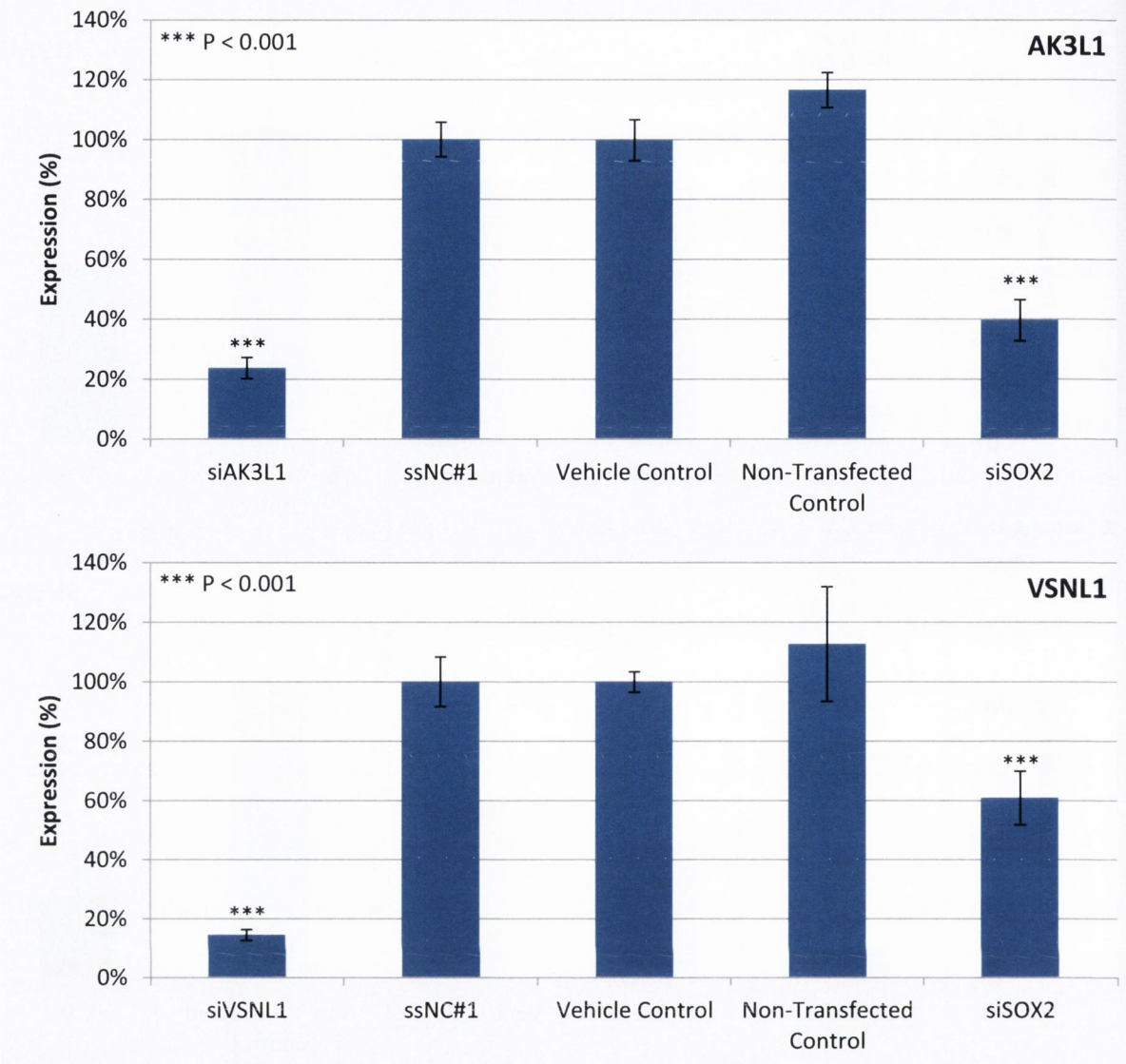


Figure 4.17—Gene expression in 2102Ep cells after transfection of siDPPA3, siSHISA2, siRBM47, siAK3L1 and siVSNL1. In all functionally transfected samples, except siRBM47 transfected cells, gene knockdown was achieved with a significant margin when compared to the ssNC#1-transfected cells (siDPPA3: P = 0.0007, siSHISA2: P = 0.0149, siRBM47: P = 0.8548, siAK3L1: P < 0.0001 and siVSNL1: P < 0.0001). The expression in the ssNC#1 samples was set at 100%. In addition, the expression of each gene was measured in cells differentiated by siSOX2 knock-down. Compared to ssNC#1-transfected cells, all genes, except RBM47, were significantly downregulated (siDPPA3: P = 0.0366, siSHISA2: P = 0.0393, siRBM47: P = 0.4424, siAK3L1: P = 0.0003 and siVSNL1: P = 0.0053). All experiments were performed in biological triplicate and the values shown are the mean expression of these triplicates ± SD. A t-test was used to compare sample means.

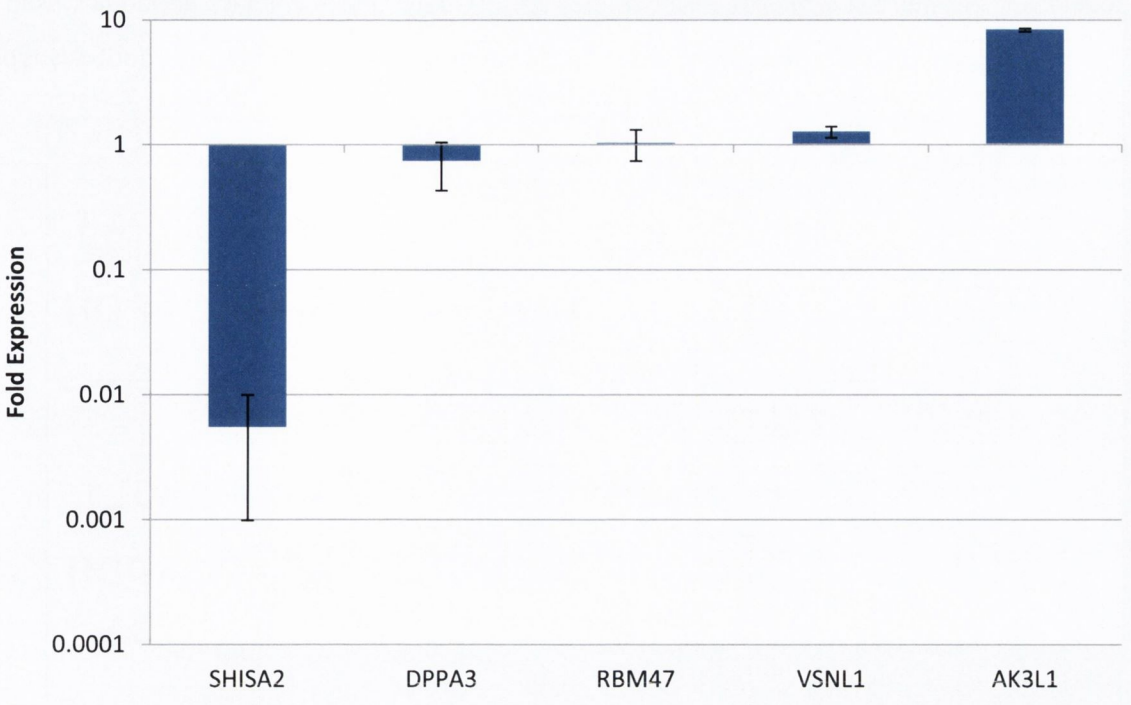


Figure 4.18—Relative expression of DPPA3, SHISA2, RBM47, AK3L1 and VSNL1 in undifferentiated 2102Ep cells. RBM47 Expression was set at 1 and the expression of all other genes were calculated relative to this. All experiments were performed in biological triplicate and the values shown are the mean expression of these triplicates ± SD.

4.4.4.3 — Protein Expression Analysis

To confirm siRNA-induced downregulation at the protein level, Western blots were optimised and conducted for the downregulated genes. Primary anti-bodies for RBM47 were tested at multiple concentrations, however the extensive presence of non-specific binding could not yield conclusive results. As knock-down of this gene was unsuccessful, further experimentation with this gene was abandoned. Western blots on DPPA3, SHISA2, AK3L1 and VSNL1 were performed by the protocol used in Section 4.3.3.3. For DPPA3 an inhibition peptide was used to determine band size. Results can be found in Figure 4.19. Downregulation of protein matches the downregulation of mRNA expression in all functionally transfected samples. For DPPA3, the inhibition peptide prevented antibody binding at 28 kDa and 18 kDa bands. The predicted molecular weight of the protein (Q6W0C5, UniProtKB/Swiss-Prot) is 17.9 kDa and no reported isoforms exist of this gene.[438] After performing a protein sequence alignment (blastp, BLAST®, NCBI), the highest scoring sequence was for hCG1987838 (accession number: EAW54683), isoform CRA_a, a predicted protein with a predicted molecular weight of 27.4 kDa.[439] This protein could account for the band at 28kDa due to the high sequence coverage of 98% (E-value = $6 \cdot 10^{-106}$) and similar band size. The closest validated gene is FAM156A (accession number: NP_054857), which has a sequence coverage of 87% (E-value = $4 \cdot 10^{-4}$) and a molecular weight of 24.4 kDa. SHISA2 and RBM47 protein expression

could not be measured due to the extreme lack of specificity of primary antibodies used (not shown), despite the extensive optimisation with multiple commercially available antibody types.

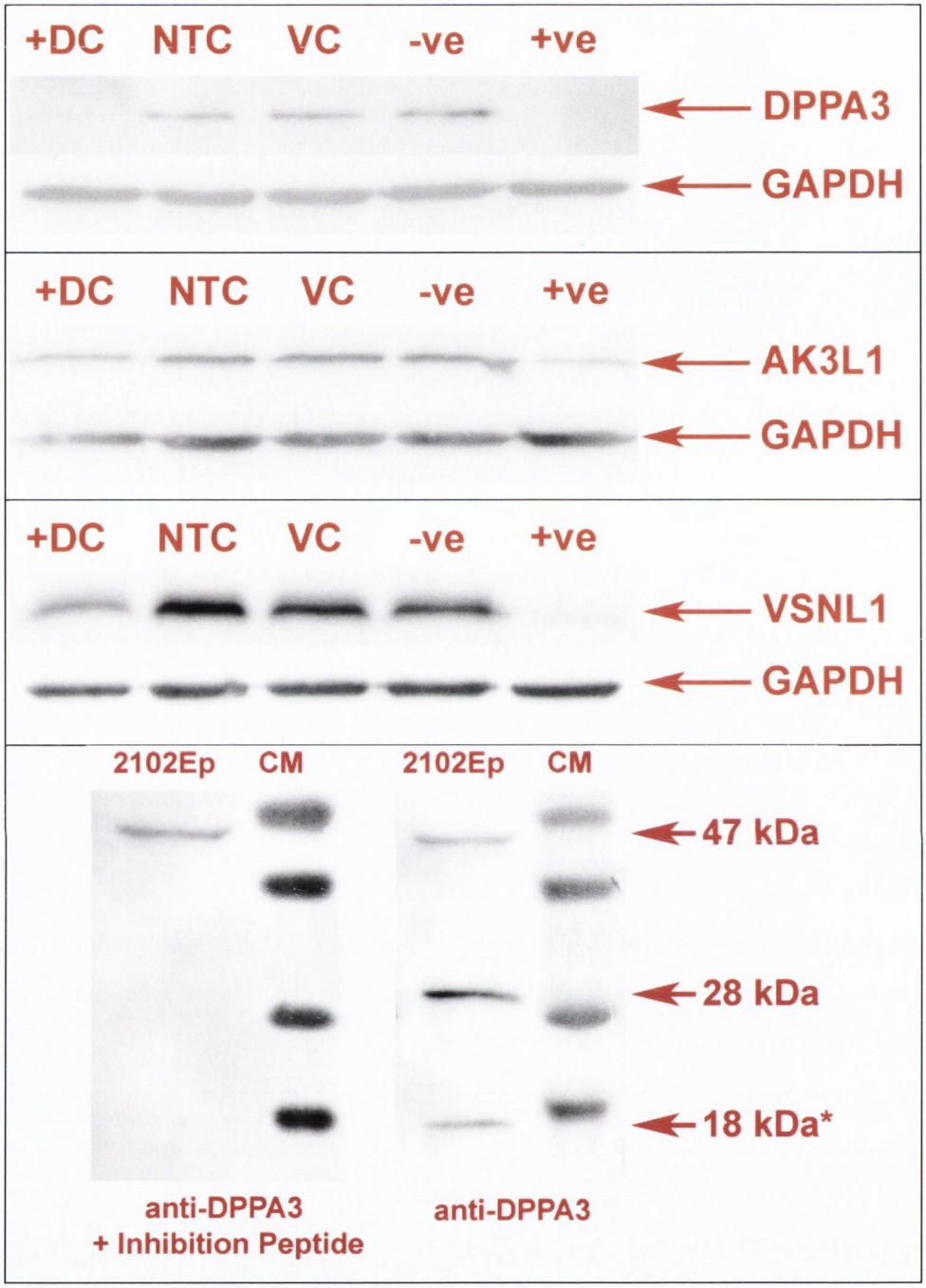


Figure 4.19—Western blots of silenced genes in 2102Ep cells. Protein from positive controls (+ve), ssNC#1 samples (-ve), vehicle controls (VC) and non-transfected controls (NTC) included. Also included were siSOX2-transfected samples as a differentiation control (+DC). The bottom image depicts blots incubated with and without a anti-DPPA3 inhibition peptide. For this protein from undifferentiated 2102Ep cells was used (2102Ep). The colour marker is also included (CM).

4.4.4.4 — Expression Analysis of Pluripotency Markers

For all samples pluripotency marker expression was measured to assess whether pluripotency was reduced and differentiation perhaps occurred. This was performed in a similar manner to Section 4.4.4.2 with the method described in Section 4.3.3.2. Results are displayed in Figure 4.20.

The results show no signficant deacrease in stemness marker expression in 2102Ep cells by the silencing of any of the succesfully targeted genes. Conversely, siAK3L1 transfected cells show a significant 1.21-fold increase in SOX2 expression and siVSNL1 transfected cells show a significant increase in SOX2 and NANOG by 1.3 and 1.8-fold. The differentiation control, siSOX2 transfected cells show a significant decrease of SOX2 and NANOG expression, but unusually, OCT4 expression remains unchanged.

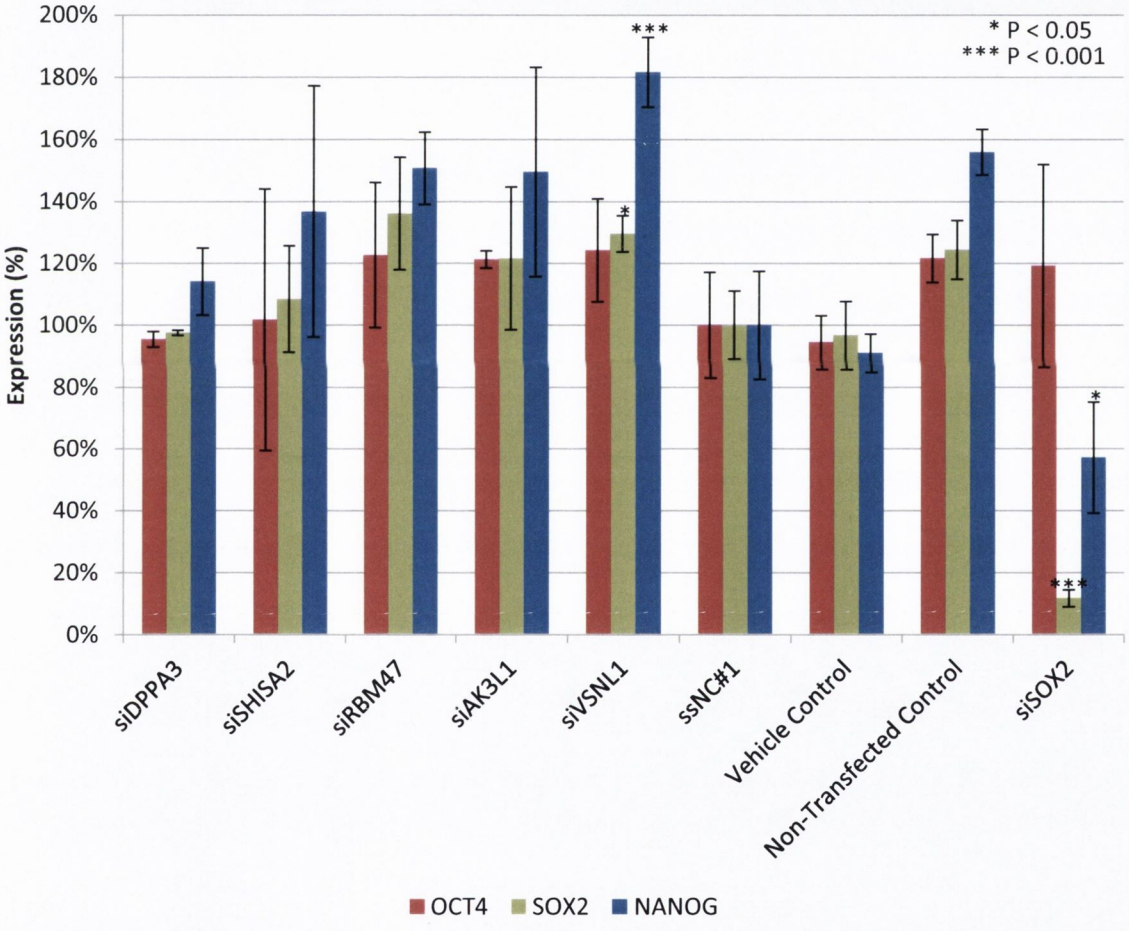


Figure 4.20—Stemness marker expression in 2102Ep cells transfected with six siRNA. When compared to ssNC#1-transfected 2102Ep cells, cells transfected with siVSNL1 showed a significant increase in SOX2 and NANOG expression (P = 0.0149 and P = 0.0007 respectively). Cells transfected with siDPPA3, siSHISA2 or siAK3L1 did not show a significant change in stemness marker expression. siSOX2 transfected cells show a significant decrease in SOX2 and NANOG expression (P = 0.0002 and P = 0.0230 respectively). The expression in the ssNC#1 samples was set at 100%. All experiments were performed in biological triplicate and the values shown are the mean expression of these triplicates ± SD. A t-test was used to compare sample means.

4.4.4.5 — Alkaline Phosphatase Assay

The enzymatic staining of AP was performed on all experimental and control samples in a similar manner to that described in Section 4.4.4.5. The protocol for this procedure can be found in Section 2.1.9. All cells were split on day three post-transfection to provide a suitable confluency for the AP staining method, which was performed on day four. Compared to the negative controls, AP staining was not visibly reduced in 2102Ep cells transfected with siDPP3, siSHISA2, siRBM47, siAK3L1 and siVDSNL1 as shown in Figure 4.21. A reduction of stain was seen in siSOX2-transfected cells, which coincides with similar results in Section 4.4.2.3 and indicates that a degree of cell differentiation is taking place.

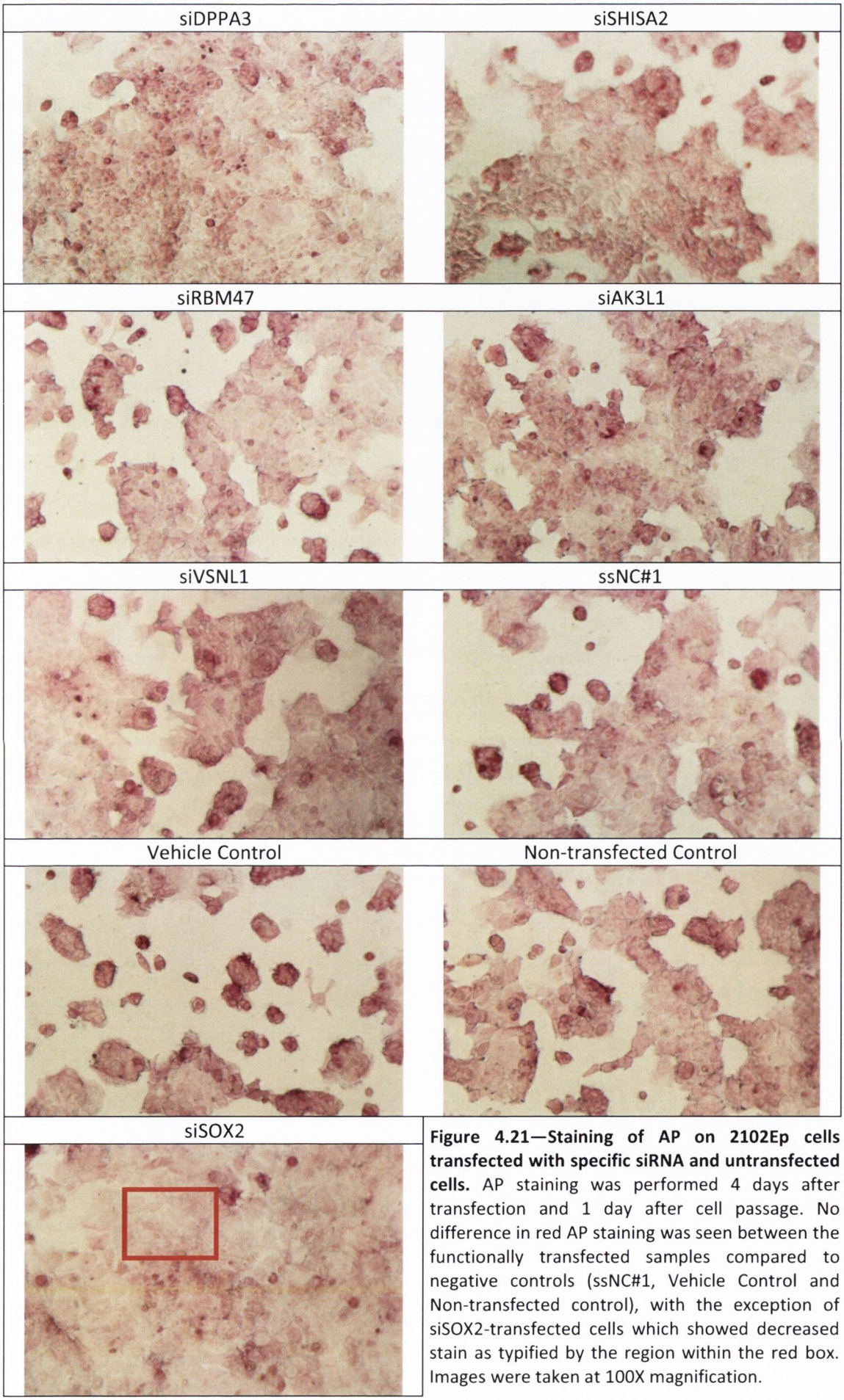


Figure 4.21—Staining of AP on 2102Ep cells transfected with specific siRNA and untransfected cells. AP staining was performed 4 days after transfection and 1 day after cell passage. No difference in red AP staining was seen between the functionally transfected samples compared to negative controls (ssNC#1, Vehicle Control and Non-transfected control), with the exception of siSOX2-transfected cells which showed decreased stain as typified by the region within the red box. Images were taken at 100X magnification.

4.4.4.6 — Assessing Cell Viability

Analysis of cell viability in 2102Ep cells was performed four days post-transfection. The most reliable period for testing cell viability is during logarithmic growth, before the cells reach confluency. At three days post-transfection 2102Ep and NTera-2 cells reach confluency, complicating accurate cell viability measuring. To resolve this issue, it was decided that cell viability would be tested at four days post-transfection with an intermediate split after three days to provide the cells with a confluency that facilitated logarithmic growth. To overcome the difficulties of splitting cells in small wells, they were seeded and transfected on 24-well plates (in triplicate) as described in Section 2.2.2.5. As cell splitting could result in an uneven distribution of cells relative to other samples, such as negative controls, the all cells were replated among two wells during splitting. For the measurement of 2102Ep cell viability, the MTT-based cell viability assay was performed on 24-well plates as described in Section 2.1.8. To measure the absorbance of the resultant formazan solution after the assay was applied, 100 µl aliquots from each well was transferred to three wells on a 96-well plate and scanned with the plate reader as used in the protocol. This method provided 2 × 6 technical replicates per original sample and 36 replicates for each control group, when biological triplicates were tested. The results are displayed in Figure 4.22. When compared to the ssNC#1-transfected cells (ssNC#1), these showed that cell viability was not significantly lower in any of the cells with a specifically silenced gene. Unexpectedly, this also holds true for SOX2-silenced cells. As the ssNC#1-transfected cells have quite high error, the cell viability of all samples was also compared to the vehicle control. Again, here no significant difference could be found.

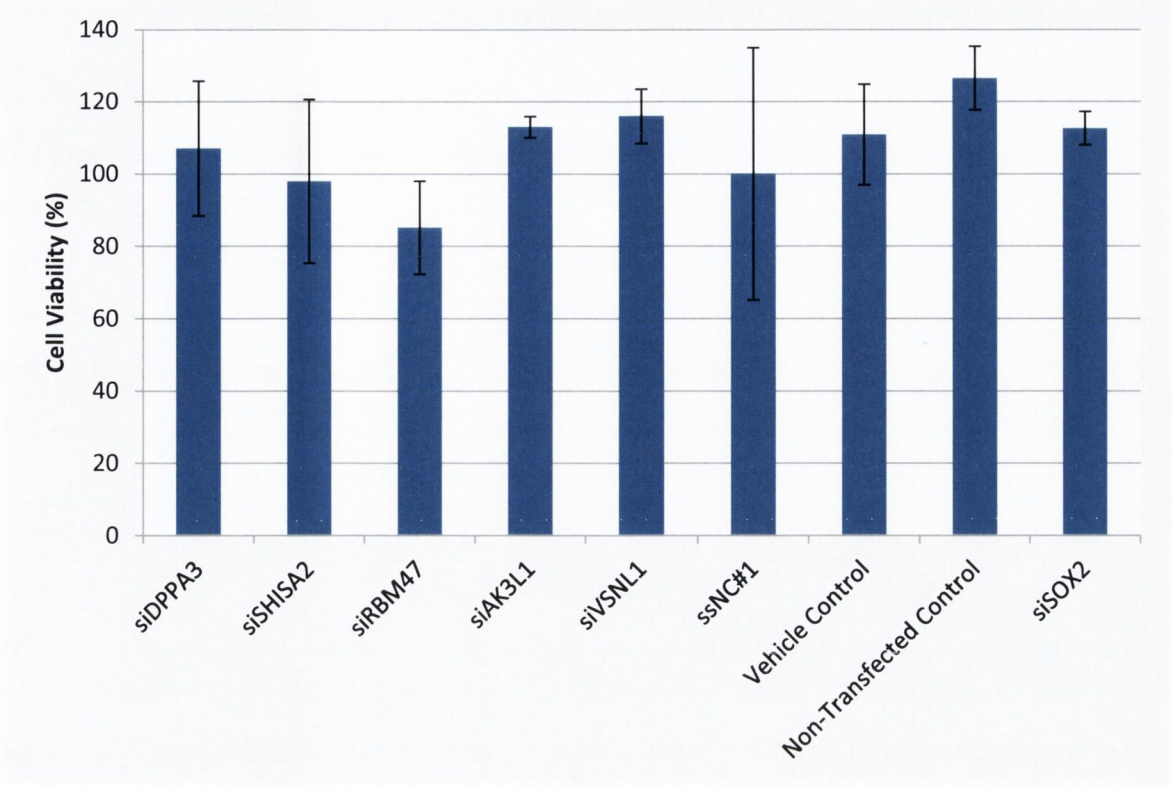


Figure 4.22—Cell viability of 2102Ep cells four days after specific gene silencing. No significant difference in cell viability could be found for experimental and control samples compared with the ssNC#1-transfected cells. When compared to the vehicle control, none of the samples were significantly different either. The cell viability of the ssNC#1-transfected cells was set at 100%. All experiments were performed in biological triplicate and the values shown are the mean cell viability of these triplicates ± SD. A t-test was used to compare sample means.

4.4.5 — Functional Silencing of Five Genes in NTera-2 Cells

To investigate the relevance of DPPA3, SHISA2, RBM47, AK3L1 and VSNL1 to NTera-2 stem cell maintenance, these genes were silenced during an experiment similar to and simultaneous with the five gene knock-down in 2102Ep cells (see Section 4.4.2.3). For this experiment the cells were incubated three after transfection, subsequently split 1:2 as described in Section 2.1.5 and incubated for another three days to a total of 6 days of post-transfection incubation as a comparison. The rest of the cells were washed in 1X PBS, pelleted and stored at -80°C. The aim of the 6-day incubation was to determine whether longer term knock-down had any phenotypic effect on hECCs, which would otherwise possibly be missed at three days.

4.4.5.1 — Phenotypic Analysis

Images were taken at four and 6 days with a light microscope. Images from cells grown for four days post-transfection are shown to visualise cell morphology and growth, while avoiding misrepresentation due to high confluence. In Figure 4.23, which showed cells incubated for four days post-transfection, lower cell density was observed in cells transfected with siDPPA3 when compared to those transfected with ssNC#1. Interestingly, siAK3L1 and siVSNL1 transfected cells appeared to have a higher density than the negative controls. At 6 days, as shown in Figure 4.24, these differences in density seemed to persist, however this was harder to directly see due to high confluence in all wells. Little difference in morphology was observed for all transfected NTera-2 cells, while cells exposed to 10^{-2} µM ATRA were visibly differentiating. Similar observations were made at three days post-transfection, although an assessment was complicated due to the high confluency of the cells.

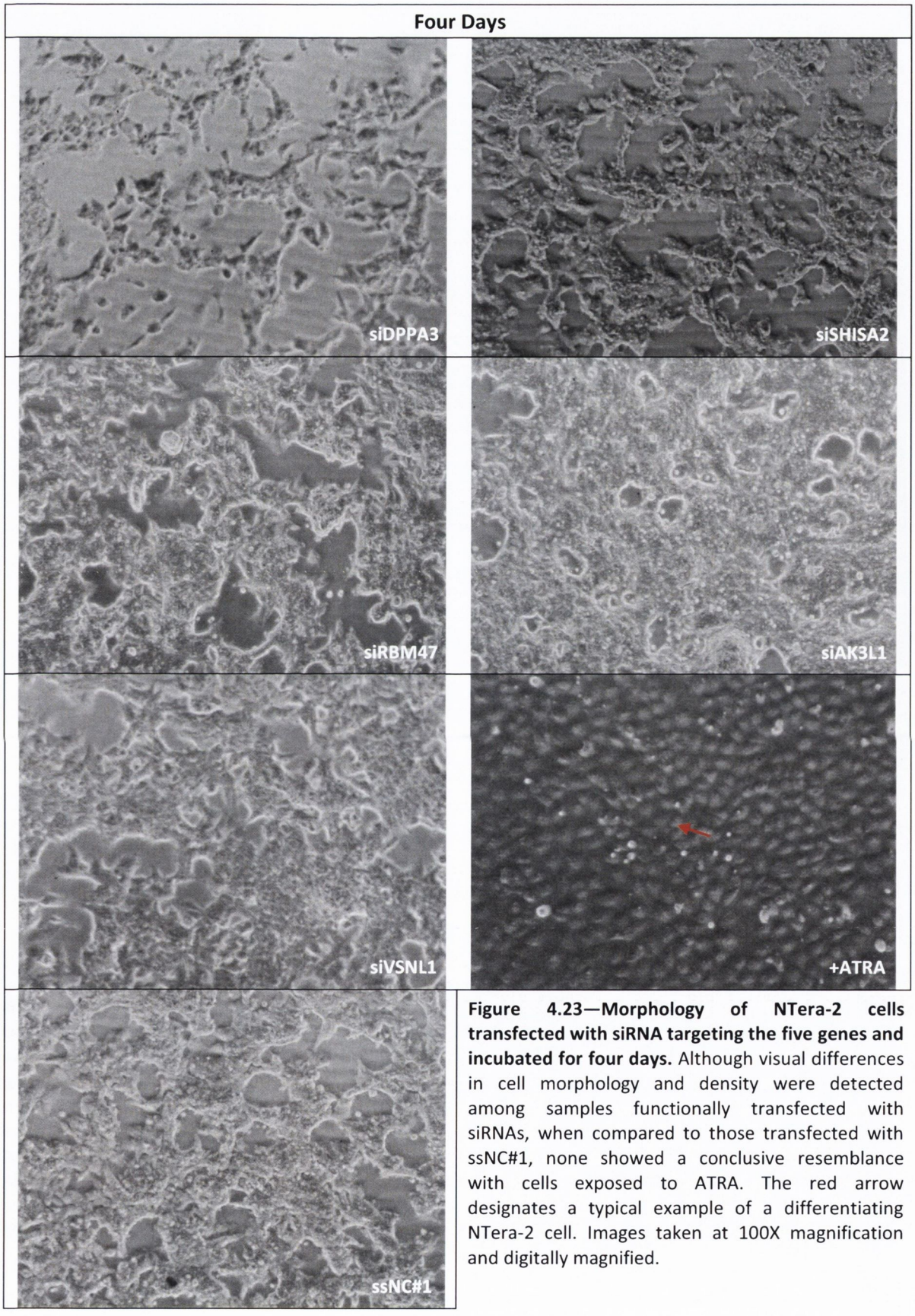


Figure 4.23—Morphology of NTera-2 cells transfected with siRNA targeting the five genes and incubated for four days. Although visual differences in cell morphology and density were detected among samples functionally transfected with siRNAs, when compared to those transfected with ssNC#1, none showed a conclusive resemblance with cells exposed to ATRA. The red arrow designates a typical example of a differentiating NTera-2 cell. Images taken at 100X magnification and digitally magnified.

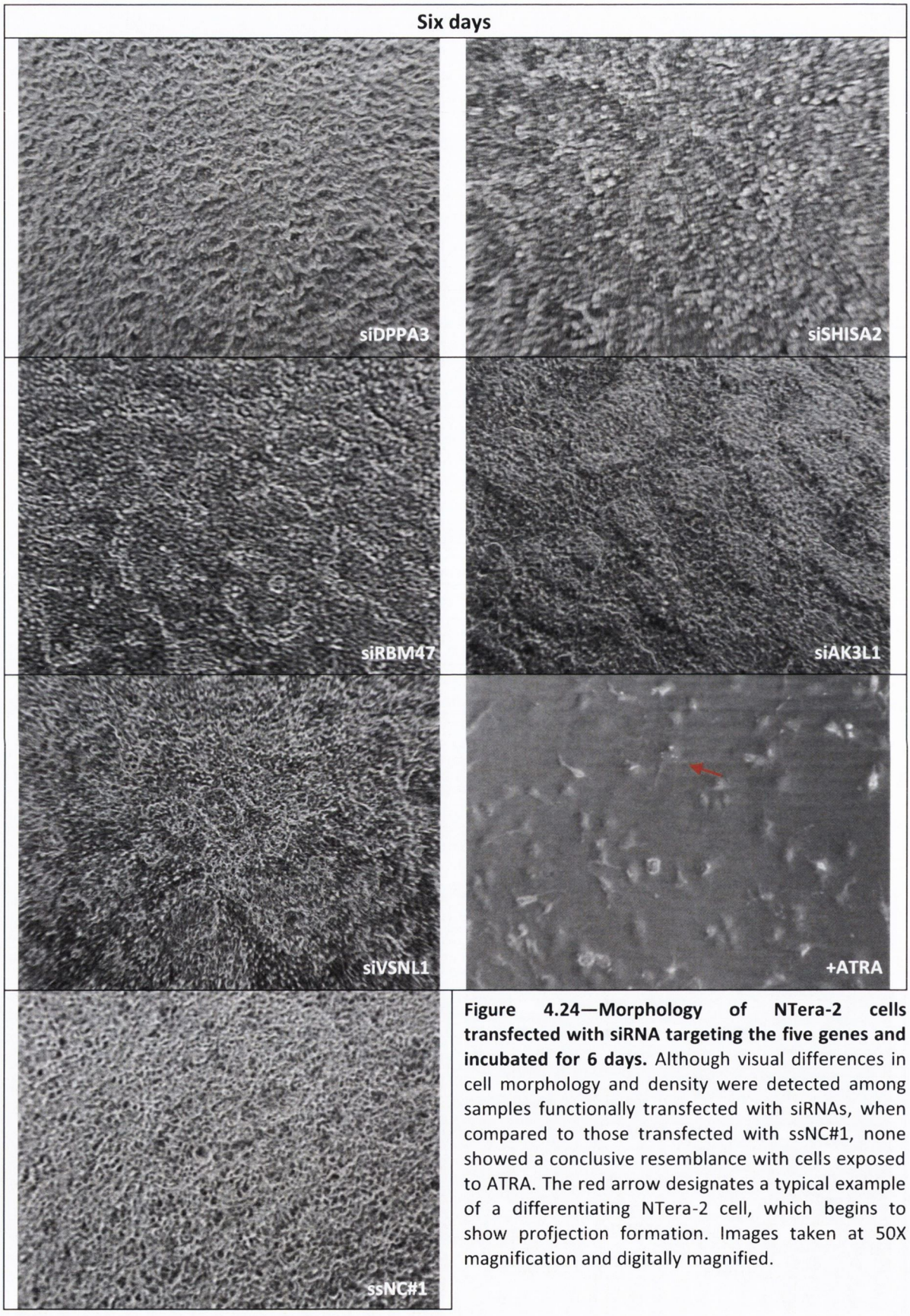
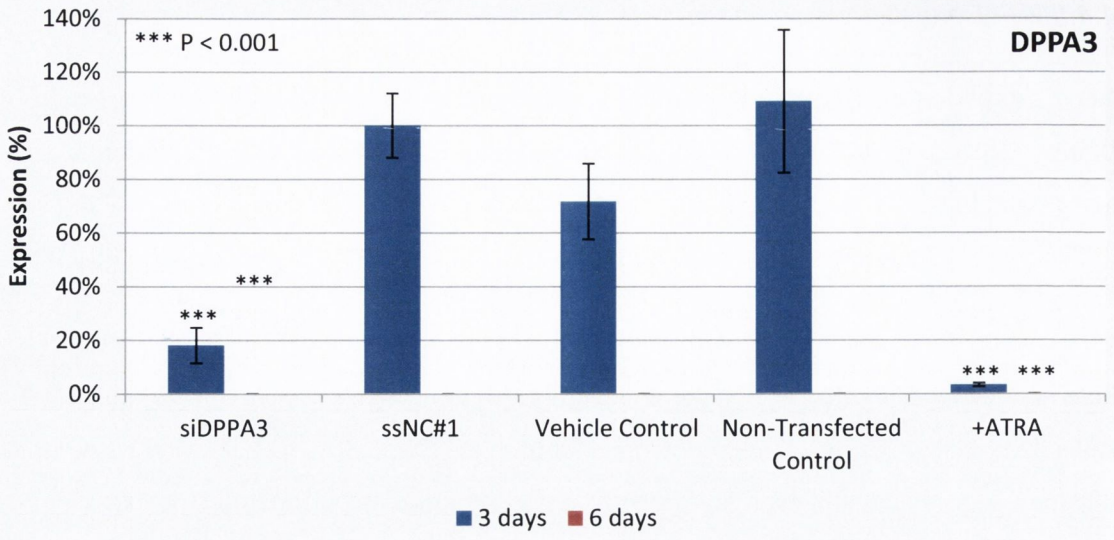


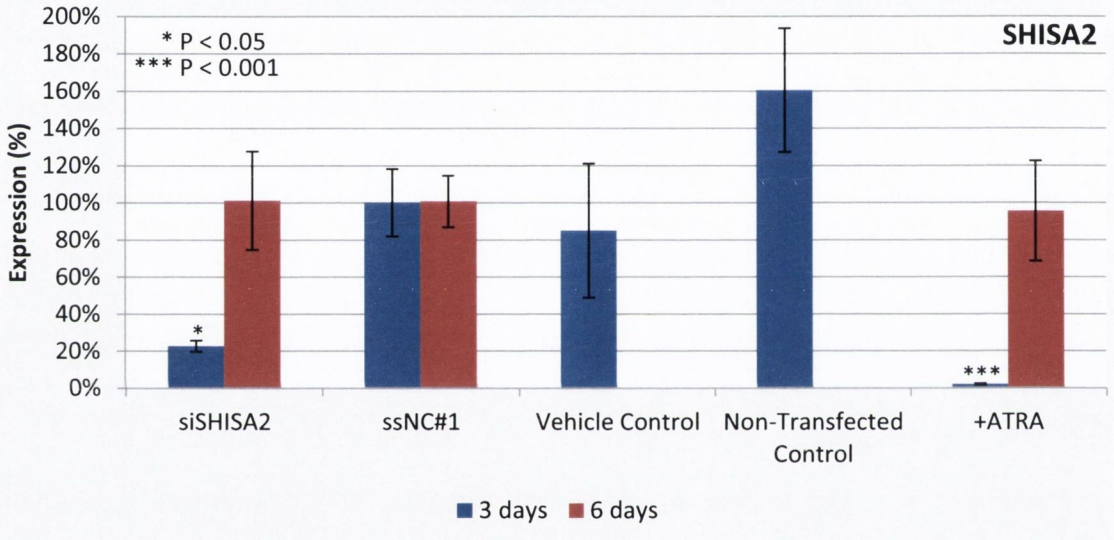
Figure 4.24—Morphology of NTera-2 cells transfected with siRNA targeting the five genes and incubated for 6 days. Although visual differences in cell morphology and density were detected among samples functionally transfected with siRNAs, when compared to those transfected with ssNC#1, none showed a conclusive resemblance with cells exposed to ATRA. The red arrow designates a typical example of a differentiating NTera-2 cell, which begins to show profjection formation. Images taken at 50X magnification and digitally magnified.

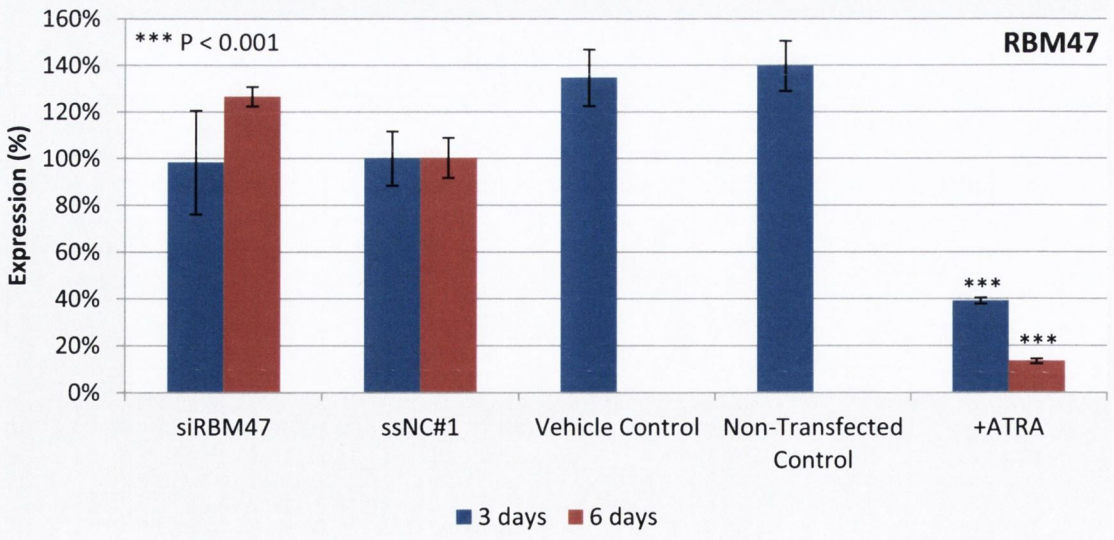
4.4.5.2 — Gene Expression Analysis

Gene expression analysis was performed as described in Section 4.3.3.2. The expression of the DPPA3, SHISA2, RBM47, AK3L1 and VSNL1 three and 6 days after siRNA transfection is displayed in Figure 4.25.

With the exception of RBM47, three days post-transfection the expression of functionally silenced genes were significantly lower with knock-down of approximately 77-94%. As in 2102Ep cells, RBM47 siRNA was ineffective for the silencing of this gene. After 6 days DPPA3 and VSNL1 were still substantially and significantly downregulated by 71% and 90% respectively, however AK3L1 regained expression and was significantly downregulated by 40% compared to the ssNC#1 samples. SHISA2 regained full expression. 10^{-2} μM ATRA significantly downregulated all genes at three days by 60-98%, and remained low except for SHISA2 which regained full expression. The expression of the five genes in the non-transfected controls ranked relative to each other can be found in Figure 4.26. Interestingly, DPPA3 expression was significantly lower in NTera-2 cells than in 2102Ep cells relative to the other genes expressed in the cells, but also in estimated absolute values. DPPA3 expression relative to GAPDH was approximately 17-fold higher in undifferentiated 2102Ep cells compared to undifferentiated NTera-2 cells(data not shown).


DPPA3
*** P < 0.001
Expression (%)
140%
120%
100%
80%
60%
40%
20%
0%
***
***
***
***
siDPPA3
ssNC#1
Vehicle Control
Non-Transfected Control
+ATRA
3 days
6 days


SHISA2
* P < 0.05
*** P < 0.001
Expression (%)
200%
180%
160%
140%
120%
100%
80%
60%
40%
20%
0%
*
***
siSHISA2
ssNC#1
Vehicle Control
Non-Transfected Control
+ATRA
3 days
6 days


RBM47
*** P < 0.001
Expression (%)
160%
140%
120%
100%
80%
60%
40%
20%
0%
***
***
siRBM47
ssNC#1
Vehicle Control
Non-Transfected Control
+ATRA
3 days
6 days

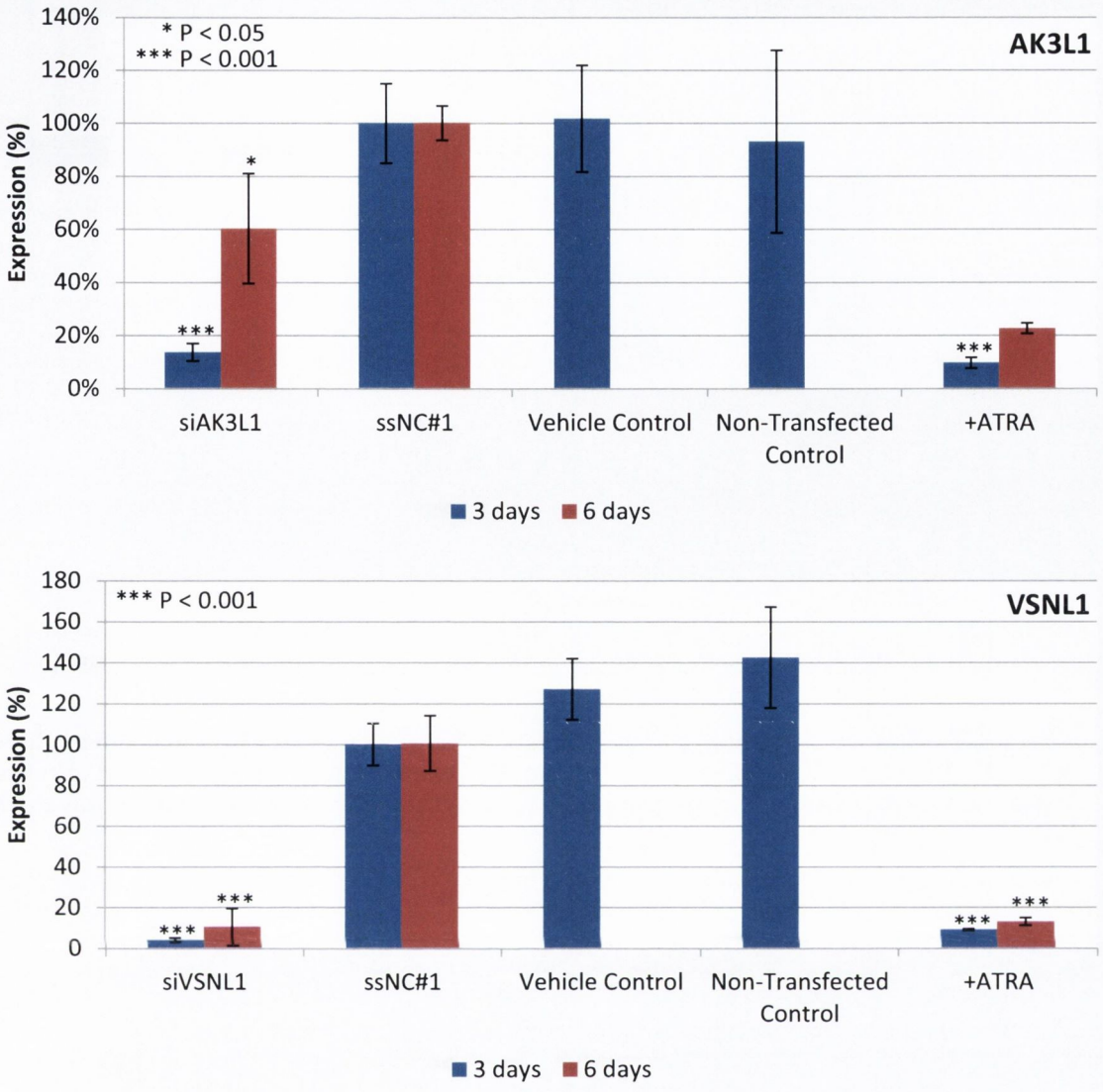


Figure 4.25—Gene expression in NTera-2 cells after transfection of siDPPA3, siSHISA2, siRBM47, siAK3L1 and siVSNL1. With the exception of RBM47, at three days post-transfection all genes were significantly downregulated in functionally transfected cells compared with those transfected with ssNC#1 (DPPA3: P = 0.0005, SHISA2: P = 0.0114, RBM47: P = 0.9134, AK3L1: P = 0.0006 and VSNL1: P < 0.0001). At six days downregulation is significant for DPPA3, AK3L1 and VSNL1 (P = 0.0013, P = 0.0336 and P = 0.0007 respectively). All genes are significantly downregulated after a three day incubation in 10^{-2} µM ATRA (DPPA3: P = 0.0002, SHISA2: P = 0.0046, RBM47: P = 0.0008, AK3L1: P = 0.0005 and VSNL1: P = 0.0001) At six days, SHISA2 expression recovered to original levels. The expression in the ssNC#1 samples was set at 100%. All experiments were performed in biological triplicate and the values shown are the mean expression of these triplicates ± SD. A t-test was used to compare sample means.

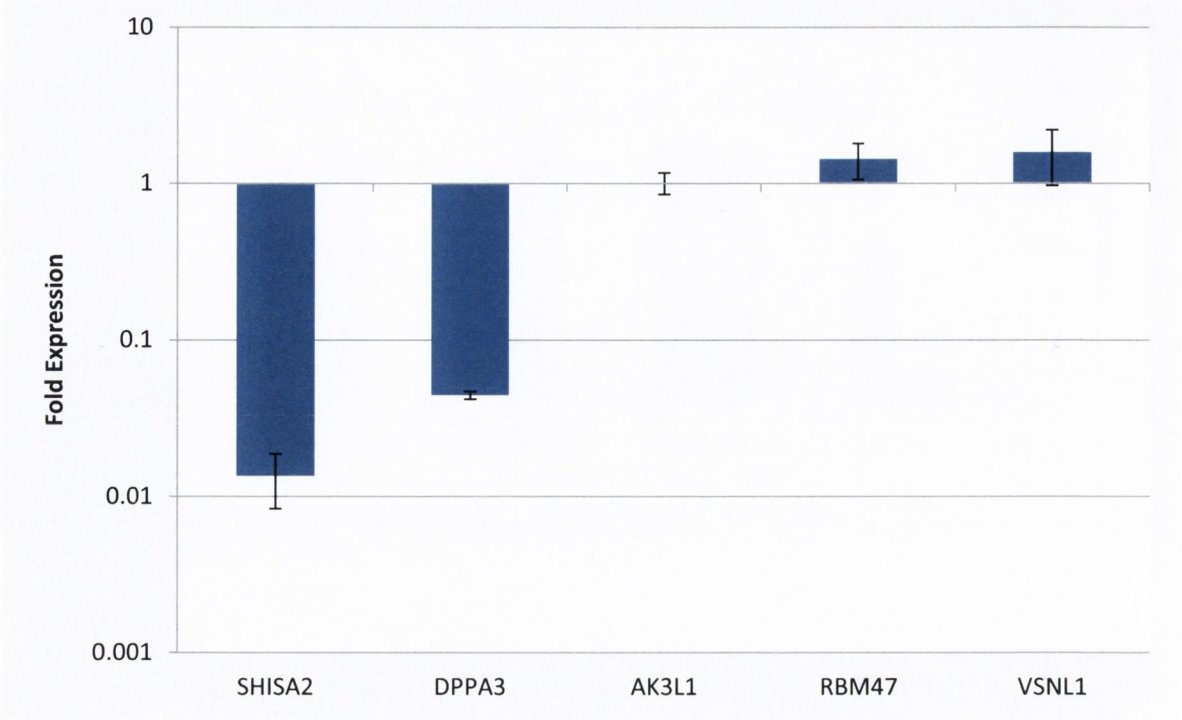


Figure 4.26—Relative expression of DPPA3, SHISA2, RBM47, AK3L1 and VSNL1 in undifferentiated NTera-2 cells. RBM47 Expression was set at 1 and the expression of all other genes were calculated relative to this. All experiments were performed in biological triplicate and the values shown are the mean expression of these triplicates ± SD.

4.4.5.3 — Protein Expression Analysis

To confirm siRNA-induced downregulation of the five genes at the protein level, antibodies were optimised and protein expression was measured by Western blot. Similar to 2102Ep experiments, primary antibodies for RBM47 could not be optimised to produce reliable results. DPPA3, SHISA2, AK3L1 and VSNL1 protein expression was measured, using the protocol used in Section 4.3.3.3. Results are shown in Figure 4.27. VSNL1 and AK3L1 protein was visibly silenced. The same DPPA3 antibody used for 2102Ep cells was optimised for protein from NTera-2 cells. In line with mRNA expression results, protein expression of DPPA3 was much lower in NTera-2 cells compared to 2102Ep cells. Unfortunately, a visible blot of DPPA3 could not be produced. A 28 kDa band, possibly hCG1987838 isoform CRA_a, did appear and its expression was blocked by the inhibition peptide as shown in the bottom image of Figure 4.27.

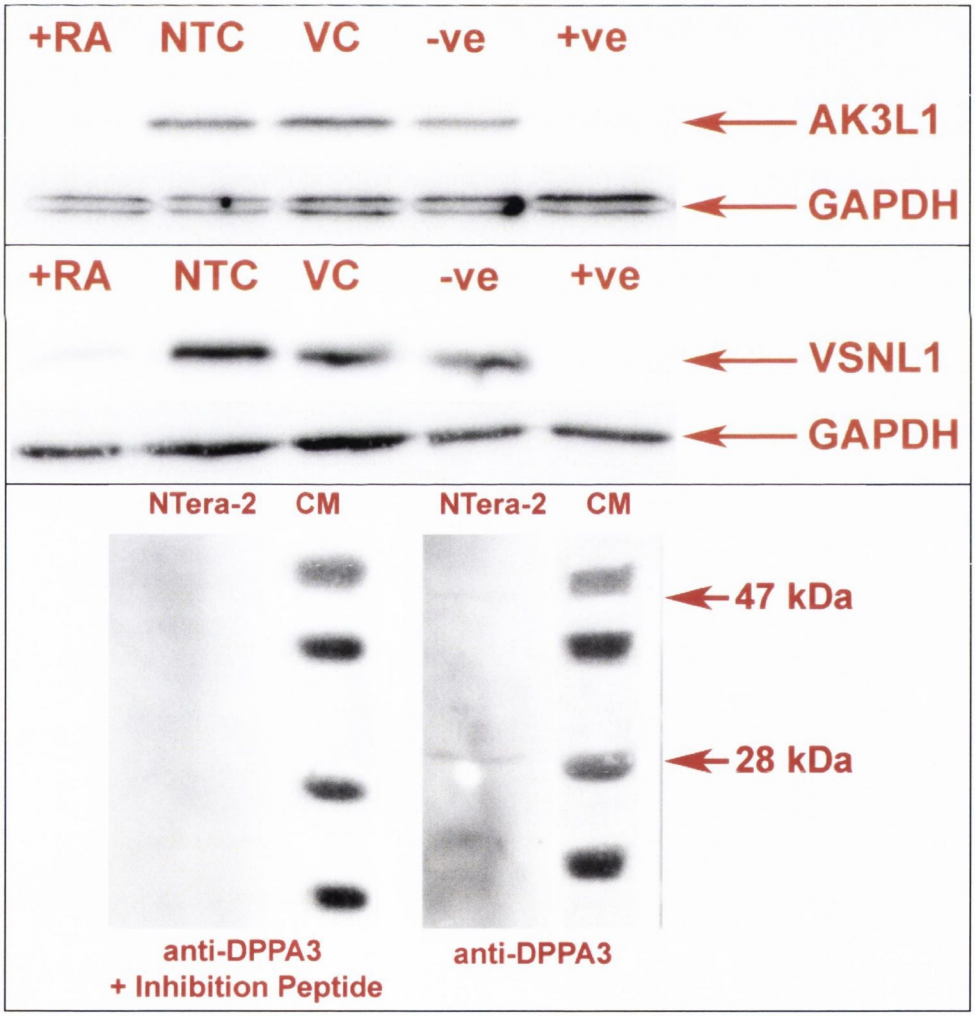


Figure 4.27—Western blots of silenced genes in NTera-2 cells. Protein from positive controls (+ve), ssNC#1-transfected cells (-ve), vehicle controls (VC) and non-transfected controls (NTC) were included. Also included were ATRA-differentiated samples as a differentiation control (+RA). The bottom image depicts blots incubated with and without a anti-DPPA3 inhibition peptide. For this, protein from undifferentiated NTera-2 cells was used (NTera-2).

4.4.5.4 — Expression Analysis of Pluripotency Markers

For all samples pluripotency marker expression was measured to assess whether pluripotency was reduced. This was performed in similar manner to Section 4.4.2.3 with the method described in Section 4.3.3.2. Results are displayed in Figure 4.28.

Although some significant differences in stemness marker expression reported for functionally transfected NTera-2 cells when compared to the ssNC#1-transfected cells, none were substantial. A significant 36% downregulation of OCT4 in siVSNL1 transfected cells was reported. This, however, was not accompanied with a simultaneous downregulation of SOX2 or NANOG and did not continue at 6 days post-transfection. SOX2 was 13% upregulated in siAK3L1 transfected cells at 6 days, however, this was not substantial and probably not biologically relevant. On the other hand, siRBM47 seemingly caused a signficant 30% increase in NANOG expression at three days post-transfection, which increased to 60% overepxression after 6 days. For cells incubated in 10^{-2} μM ATRA, OCT4 and NANOG were substantially downregulated by 74% at three days and these genes were almost completely silenced at 6 days. SOX2 expression, on the other hand, experienced no significant change at three days and was 10% downregulated at 6 days.

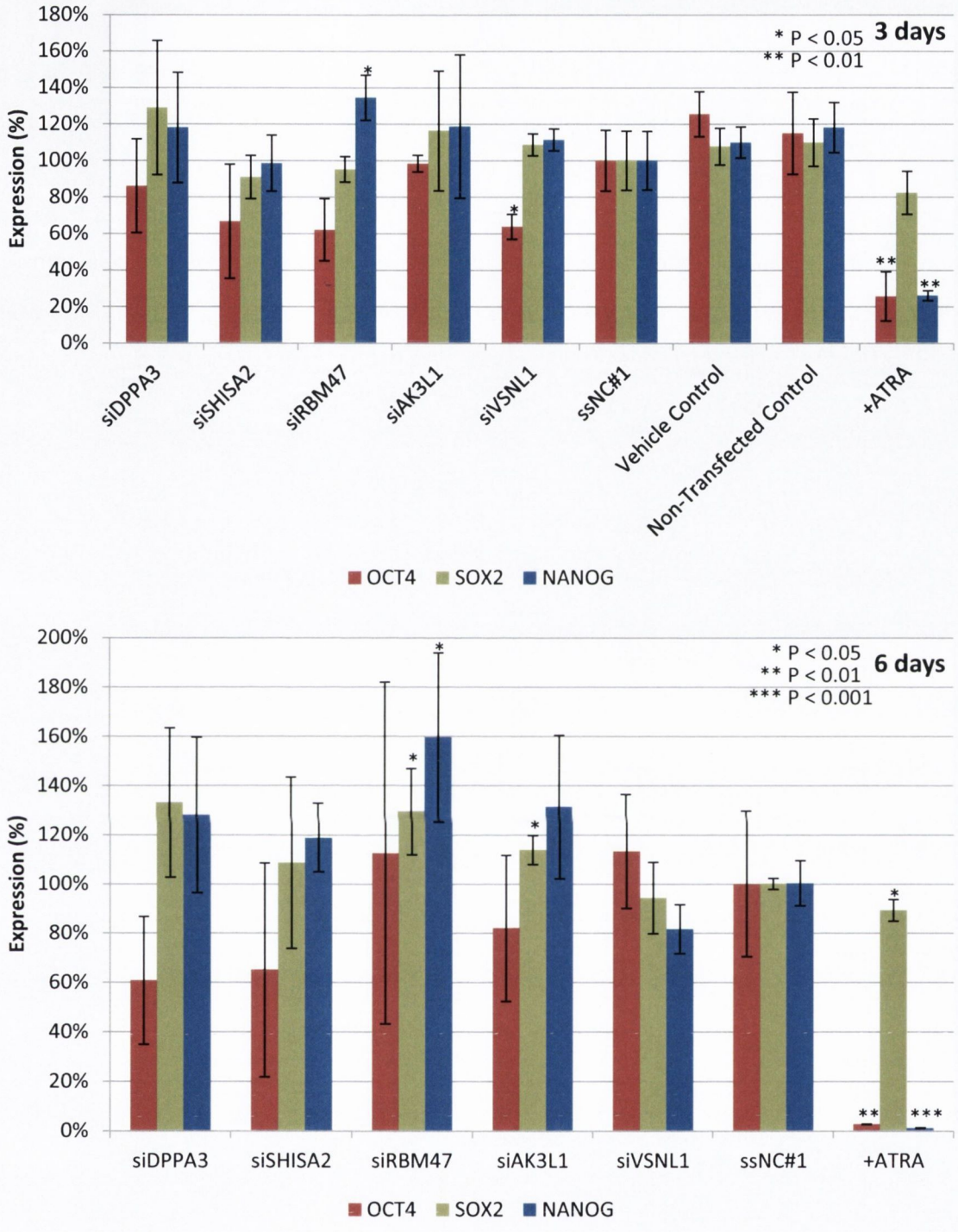


Figure 4.28—Stemness Marker expression in functionally transfected NTera-2 cells three and six days post-transfection. At three days OCT4 was significantly downregulated by 38% in siVSNL1 transfected cells when compared to ssNC#1-transfected cells (P = 0.0253). OCT4 expression did not significantly change after 6 days. SOX2 was slightly, but significantly upregulated in siRBM47 and siAK3L1 transfected cells, as was NANOG in siRBM47 transfected cells at both three and 6 days. OCT4 was significantly downregulated in cells incubated in 10^{-2} μM ATRA at both time points (3 days: P = 0.0039 and 6 days: P = 0.0042). NANOG is significantly downregulated in the same cells at both time points (3 days: P = 0.0016 and P < 0.0001), however SOX2 is only significantly downregulated by an insubstantial 11% at 6 days (P = 0.02). The expression in the ssNC#1 samples was set at 100%. All experiments were performed in biological triplicate and the values shown are the mean expression of these triplicates ± SD. A t-test was used to compare sample means.

4.4.5.5 — Alkaline Phosphatase Assay

The enzymatic staining of cell surface AP was performed on all cell samples at four days post-transfection in a similar manner as was performed in Section 4.4.4.5. As with 2102Ep cells, the NTera-2 cells were split three days after transfection and left to adhere for one day before the AP staining assay was applied. The protocol for AP staining can be found in Section 2.1.9. As can be seen in Figure 4.29, little reduction of cell surface AP expression is observed on cells transfected with specific siRNA, compared to the negative control samples. For cells that have been exposed to 10^{-2} ATRA, AP expression is visibly eliminated.

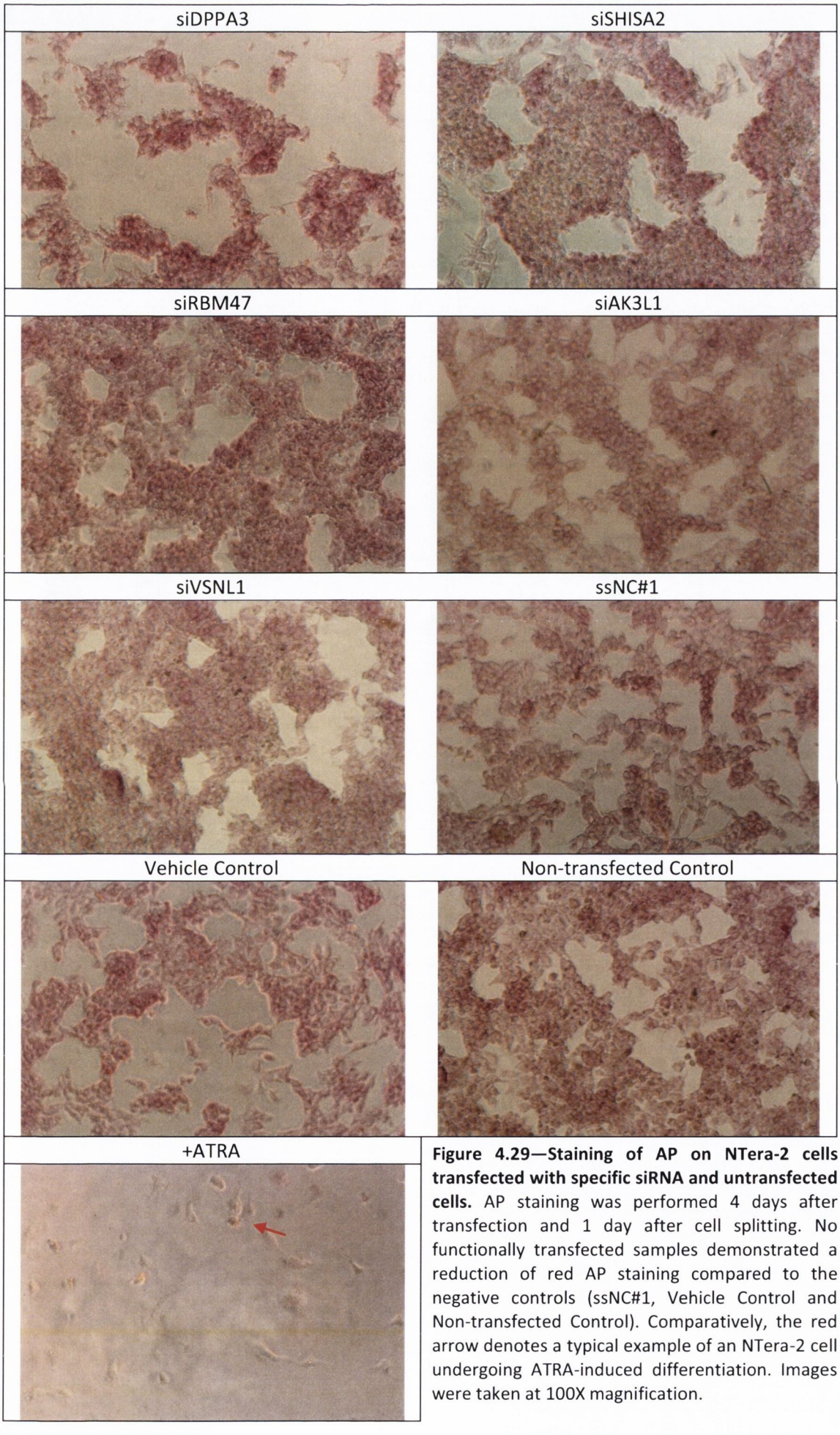


Figure 4.29—Staining of AP on NTera-2 cells transfected with specific siRNA and untransfected cells. AP staining was performed 4 days after transfection and 1 day after cell splitting. No functionally transfected samples demonstrated a reduction of red AP staining compared to the negative controls (ssNC#1, Vehicle Control and Non-transfected Control). Comparatively, the red arrow denotes a typical example of an NTera-2 cell undergoing ATRA-induced differentiation. Images were taken at 100X magnification.

4.4.5.6 — Assessing Cell Viability

Cell viability analysis of NTera-2 cells was performed in a similar manner to the method used in Section 4.4.4.6. Cell viability was measured at four days after cell transfection using the MTT cell proliferation assay as described in Section 2.1.8. These experiments showed no significant difference between the functional siRNA-transfected NTera-2 cells and the negative controls (ssNC#1-transfected cells, vehicle controls and non-transfected controls).

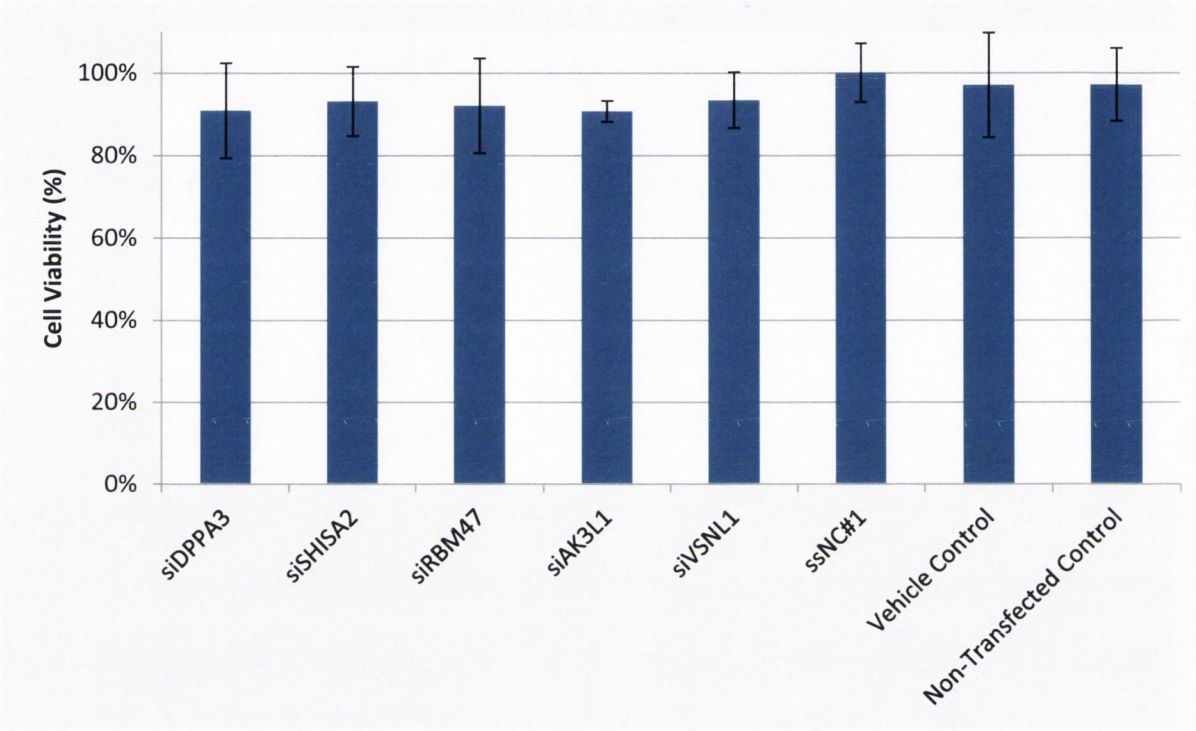


Figure 4.30—Cell viability of 2102Ep cells four days after specific gene silencing. Cell viability of experimental samples was not significantly different from those transfected with ssNC#1. The cell viability of the ssNC#1-transfected cells was set at 100%. All experiments were performed in biological triplicate and the values shown are the mean cell viability of these triplicates ± SD. A t-test was used to compare sample means.

4.5 — Discussion

The pluripotent stem cell phenotype is regulated by a core set of transcription factors, which cooperatively integrate in and interact with a large network of regulatory proteins and non-coding genes. This network is responsive to the intrinsic program of the cell, but it is also crucially dependent on extrinsic signals from a protected microenvironment, sometimes called the 'stem cell niche'.[68] The tumourgenicity of cells can arise from a interchange of intrinsic cellular alterations (e.g. genome mutations, chromosomal abnormalities and epigenetic changes) and dynamic microenvironmental conditions (e.g. hypoxia, acidic stress or the presence or absence of particular extracellular signalling molecules).[68,114,119,457] As described in Chapter 1, it is believed that these alterations give rise to a population of cells with an aberrant stem cell-like phenotype that promotes the overall tumourgenicity, proliferation and maintenance of cancer. A substantial amount of effort is invested to find ways to remove this phenotype from cancer cells in a manner that is least impacting to normal cell homeostasis. Many approaches, aimed to target the intrinsic and external factors are continually suggested and some have been implemented with varied success (see Section 1.4 for more details on this). This is difficult because of the plastic and adaptable nature of the stem cell phenotype, but also because of the similarity of stem cell-like cancer cells to normal, non-cancerous (stem) cells.

One of the approaches, adopted by some laboratories, is the targeting of genes believed to be responsible for contributing to stemness in cancer, ideally without affecting the function of normal cells. This approach is discussed in the introduction of this chapter and the work performed in this chapter aimed to contribute to this research by selecting gene targets which could potentially have a role in maintaining the stem cell phenotype of embryonal carcinoma. The importance of these genes for hECC stemness was tested by silencing them in 2102Ep and NTera-2 cells with the aim to induce differentiation in these cells. Five of these targets were selected on the basis of their high expression change during early ATRA-induced differentiation of NTera-2 cells, as performed by a previous study from this laboratory. The aim was to find genes that can possibly differentiate hECCs exclusively, with a low hypothesised chance of affecting their non-malignant counterparts, hESCs. In addition to this, SOX2, a crucial pluripotency marker, was silenced in the apparently nullipotent and retinoic acid resistant 2102Ep cells to assess their capacity to differentiate and to test whether the method of gene silencing used could indeed be used to perform this in this cell line, but also to search for novel findings associated with SOX2 function in ECCs.

4.5.1 — Successful Establishment of a Laboratory Protocol for Transient Gene Silencing in 2102Ep and NTera-2 Cell Lines

The silencing of genes in 2102Ep and NTera-2 cells was conducted by RNAi, with manufacturer-designed, propriety siRNAs from Ambion, which were delivered into the cells by cationic lipid transfection. Prior to experimental silencing of genes, siRNA delivery and gene silencing was optimised for 2102Ep and NTera-2 cell lines. With the transfection agent, Lipofectamine™ 2000, transfection of siRNAs into the 2102Ep cells was optimised with an siRNA against SOX2. In NTera-2 cells, gene silencing optimisation was carried out together with fellow student, Salah ElBaruni, and optimisation was performed with an siRNA against GAPDH, delivered by the transfection agent, Lipofectamine™ RNAiMAX™.

After optimisation with various transfection agent and siRNA concentrations, successful and efficient gene knock-down was achieved with little remaining expression of the transfection optimisation control genes remained when compared to the non-targeting controls (30% after 24 hours in 2102Ep cells and 12% after 48 hours in NTera-2 cells). Protein expression, measured for only for SOX2 in 2102Ep cells, was also reduced beyond detection, while cytotoxicity from the transfection process to both cell lines was relatively low.

4.5.2 — Experimental Silencing of SOX2 Induces Differentiation in 2102Ep and NTera-2 Cell Lines

The experimental silencing of SOX2 in 2102Ep and NTera-2 cells was conducted with the optimal method used for siRNA transfection. This resulted in a very strong downregulation of SOX2 in these cells as can be seen in Sections 4.4.2.2 and 4.4.3.2. The morphology of both cell lines changed radically and in 2102Ep cells the expression of OCT4 and NANOG, besides that of SOX2 was reduced significantly after three days. The pluripotency marker expression pattern in NTera-2 cells differed from 2102Ep cells. The downregulation of SOX2 after three days was less strong, but the biological significance of its reduction was confirmed by the substantial reduction of NANOG expression at the same time point. Although OCT4 expression did not significantly change after three days, at six days after siSOX2 transfection it was reduced by more than half. This observation, together with the very low expression of both SOX2 and NANOG at six days, confirmed the functional success of SOX2 silencing. The discrepancy of SOX2 expression between 2102Ep and NTera-2 cells at three days could be explained by the results from a previous study that showed that SOX2 expression was approximately 11-fold higher in undifferentiated NTera-2 cells compared to undifferentiated 2102Ep cells.[280] The effectiveness of SOX2 silencing in NTera-2 cells could be comparatively affected due to this gene's much higher expression.

In 2102Ep cells, differentiation marker expression was also measured. This showed a substantial increase in the expression of NCAM1 and GATA6. These results suggested 2102Ep cell differentiation was possibly occurring towards the ectoderm and endoderm.[383] Furthermore, in both cell lines cell surface AP activity was substantially reduced after four days.

Taken together, these results strongly support the hypothesis that SOX2 is crucial to the maintenance of the 2102Ep and NTera-2 undifferentiated stem cell state and that the downregulation of this gene results in the rapid differentiation of these hECC lines.

In a previous study, SOX2 was silenced in a different hECC line, NCCIT.[378] This study showed that the siRNA-induced silencing of SOX2 resulted, besides the silencing of SOX2 mRNA and protein, also in the almost complete silencing of OCT4 and NANOG mRNA after 2-days in this cell line and a complete silencing of NANOG protein, while OCT4 protein was partially silenced. SOX2 silencing also produced cells with a visibly different morphology after 1-2 days in NCCIT cells. Similar morphological change was observed in 2102Ep and NTera-2 cells, albeit at later time points. Also, the transient differentiation profiles of pluripotency markers in 2102Ep and NTera-2 cells differed with that of NCCIT cells and with each other. This could indicate that the OCT4/SOX2/NANOG pluripotency network is differently regulated in different hECC lines. This concept is supported by the previously mentioned study, which showed, in a parallel analysis of the three hECC lines, a clear difference between the differentiation states of 2102Ep cells (trophectodermal differentiation), NTera-2 cells (incomplete differentiation) and NCCIT cells (no differentiation) upon the silencing of NANOG.[378] Further parallel time course analyses of these cell lines could elucidate the more detailed differences in the core pluripotency network.

To further elucidate the relevance of SOX2 for 2102Ep and NTera-2 cell function and its contribution to the biology of different ECC lines and CSCs, the transcriptional network of these gene was analysed in Chapter 5.

4.5.3 — Silencing of Five Selected Genes Did Not Differentiate 2102Ep and NTera-2 Cells

As discussed in Section 1.4, one of the obstacles of CSC-targeted therapy is the common difficulty with distinguishing the CSC phenotype from the often similar phenotype of NSCs. Genes that distinguish CSCs from NSCs are important finds, as they can be targeted for therapy. As CSCs of teratocarcinomas, ECC lines such as 2102Ep and NTera-2 were postulated to be useful CSC models for finding CSC-specific genes as potential therapeutic targets of differentiation therapy.

Five genes were selected for induced silencing in undifferentiated 2102Ep and NTera-2 CSC models. These genes were DPPA3, SHISA2, RBM47, AK3L1 and VSNL1 and were chosen on the

supposition that they potentially could be exclusive components of differentiation program in hECCs. The exclusivity of these genes as components of early hECC differentiation was predicted on evidence that these genes were not a part of differentiation of hESCs. As described in Section 1.13 and supported by previous research, hECCs, and particularly the 2102Ep cell line, phenotypically correspond closely to hESCs, in part because of the high expression of core pluripotency genes in both cell lines.[280,318,440] The genes selected for silencing in this chapter were all highly downregulated in NTera-2 cells after three days of exposure to 10^{-2} mM ATRA. These genes were considered not to be crucial for the hESC phenotype and did not differentially express in two differentiating hESC lines as determined by two previous studies.[335,440] In 2102Ep and NTera-2 cells, siRNAs against the five genes listed was were transfected and cells were left to grow for three days. NTera-2 cells were incubated for an additional three days to observe any potential longer term changes.

2102Ep and NTera-2 cell morphology was not significantly altered after three days by the transfection of the specific siRNA compared to ssNC#1-transfected cells. From the five genes tested, four were successfully silenced at the mRNA level in both 2102Ep and NTera-2 cells as seen in Section 4.4.4.2 and 4.4.5.2 respectively. In NTera-2 cells, the expression of SHISA2 fully and AK3L1 partially recovered after six days. At the protein level, AK3L1, VSNL1 and DPPA3 were proven to be silenced in 2102Ep cells as shown in Section 4.4.4.3. In NTera-2 cells, AK3L1 and VSNL1 were shown to be successfully downregulated at the protein level (see Section 4.4.5.3). Absolute expression of DPPA3 has previously been found to be much lower in NTera-2 cells compared to 2102Ep cells, which made probing of DPPA3 protein by the applied anti-DPPA3 antibody prohibitively difficult in NTera-2 cells.[280] Unfortunately, at the time of this experiment, more sensitive antibodies were not available for DPPA3. SHISA2 mRNA was substantially knocked down in both 2102Ep and NTera-2 cells. However, SHISA2 protein could not be measured due to irregular and non-specific binding of available antibodies. The siRNA against RBM47 failed to silence this gene at the mRNA level in both cell lines, as shown in Sections 4.4.4.2 and 4.4.5.2. RBM47 is the least annotated gene in the set of genes tested in this experiment. No functional studies could be found for it, thus no previous use of siRNA-mediated silencing of this gene was reported. Unfortunately, the manufacturer did not provide siRNAs targeting different regions of the RBM47 mRNA, so no further optimisation was performed.

The significant and substantial downregulation of all five genes was confirmed in NTera-2 cells exposed to 10^{-2} mM ATRA for three days. With the exception of SHISA2, the downregulation of the four other genes was maintained at six days. Interestingly, SHISA2 expression was restored fully after six days of ATRA exposure. This indicates a temporal expression mechanism of SHISA2

during early NTera-2 cell differentiation, which could possibly suggest a specific, transient role during differentiation program. In 2102Ep cells, the silencing of SOX2 resulted in a partial significant downregulation of DPPA3, SHISA2, AK3L1 and VSNL1. The level of downregulation varied between 31-60% and could possibly indicate a direct regulatory relationship between the pluripotency markers and these genes. RBM47 expression was not affected in 2102Ep cells after SOX2 silencing, while it was significantly and substantially downregulated in NTera-2 cells exposed to ATRA. This could indicate that RBM47 regulation more closely linked to the retinoid pathway, rather than the core pluripotency network.

As can be seen in Sections 4.4.4.4 and 4.4.5.4, the knock down of DPPA3, SHISA2, AK3L1 and VSNL1 did not substantially affect pluripotency marker expression in 2102Ep or NTera-2 cells three days after transfection. In NTera-2 cells, the transfection of siRBM47 did seemingly induce a 60% NANOG overexpression six days after transfection, but as RBM47 was not silenced this could have been possibly an off-target effect from the siRNA. Cell surface AP expression did not change in the experimental control cells compared to the ssNC#1-transsfected cells in either cell line.

The cell viability of 2102Ep and NTera-2 cells was assessed four days after the transfection of specific siRNAs (see Sections 4.4.4.6 and 4.4.5.6). This did not show any significant difference in cell viability in either cell line. Unexpectedly, 2102Ep cell viability was not significantly altered four days after the induced silencing of SOX2. It is possible that SOX2 knock-down does not produce an immediate effect on cell viability, but that, when the cells are given more time to differentiate, self-renewal and proliferation reduce. Long-term differentiation experiments with SOX2 silencing in the hECC lines were not performed for this thesis, but could be an interesting endeavour for future research.

Overall, the lack of change of cell phenotype, differential expression of pluripotency markers and AP expression in 2102Ep and NTera-2 resulted in the conclusion that DPPA3, SHISA2, AK3L1 and VSNL1 were not crucial for maintaining the hEC phenotype. However, this verdict is still inconclusive for RBM47 and it is possible that with successful silencing of RBM47, this gene might still be proven to have a role in hEC maintenance.

In summary:

- The silencing of SOX2 in 2102Ep and NTera-2 cells resulted in the differentiation of both.
- The silencing of none of the five selected genes resulted in differentiation.

CHAPTER FIVE

GLOBAL GENE AND MICRORNA EXPRESSION ANALYSIS IN EMBRYONAL CARCINOMA

5 — Global Gene and MicroRNA Expression Analysis in Embryonal Carcinoma

5.1 — Introduction

Core pluripotency factors are of crucial importance to the maintenance of pluripotency in stem cells and ECCs. Arguably, this core consists of the three transcription factors, OCT4, SOX2 and NANOG.[78,80] This core network has been extensively investigated, and their target genes have been determined in hESCs and mouse ESCs by ChIP.[79,458] These ChIP studies identified the OCT4-centric module in ESCs, which includes, besides OCT4, SOX2 and NANOG, also SMAD1, STAT3 and TCF3, which are controlled by the upstream pathways effected by BMP, LIF and Wnt respectively.[80] Many other transcription factors are directly linked to the OCT4-centric module and within this network, OCT4 has been proposed as the main anchor point for multiprotein transcription complexes which maintain or transform the pluripotent phenotype of ESCs and ECCs.[80] Many of the transcription factors that constitute or are linked to core pluripotency regulate their own expression as the expression of each other to maintain a balanced expression rewrite.[80] OCT4 and SOX2 heterodimerise for the transcription of genes containing *cis*-acting elements in their promoter regions.[79,459] While OCT4 (and NANOG) are well characterised, the functions of SOX2 have only recently been elucidated. Since the crucial role of SOX2 in the establishment of iPSCs has been defined, the research in the function of this gene has been greatly accelerated.[81] SOX2 is a member of SRY-related HMG-box (SOX) family of transcription factors. It is believed that one of the primary functions of SOX2 in ESCs is to regulate the expression of OCT4 and therefore stabilise ESCs in their pluripotent state.[460] The loss of SOX2 results in the differentiation of hESCs, indicating it to be crucial to the maintenance of the undifferentiated state of these cells. The additional functions of SOX2 in the CNS and for neural development in the embryonic brain have also recently been elucidated.[461,462] This shows that the expression of SOX2, as regulated by STAT3, is elemental for commitment to the neural progenitor cell fate.[463] In addition, SOX2 has been detected in adult neural stem cells, indicating important functional roles beyond the embryonal state.[462] It has been shown that along with SOX1 and SOX3, SOX2 maintains neural stem cells in the embryonic and adult brain.[464]

A ChIP study to detect which genes are bound by SOX2, OCT4 and NANOG in hESCs showed that SOX2 was bound to the promoters of 7% of known protein-coding genes.[79] In this study, approximately a third of these genes were co-occupied by OCT4 and SOX2, but another third of all genes occupied by SOX2, were not occupied by either OCT4 or NANOG. This suggests SOX2 has numerous transcriptional functions in ESCs beyond the direct cotranscription influence from OCT4 or NANOG.

SOX2, often together with OCT4 and NANOG, has been shown to regulate miRNA expression. In the ChIP study from the last paragraph, SOX2 was shown to be bound to the promoter region of host genes of several intronic miRNA (miR-7-1, miR-22, miR-32, miR-135b, miR-137, miR-204, miR-205 and miR-301).[79] This study only included protein-coding genes and as such did not report the association of SOX2 to non-coding loci, such as intergenic miRNA. Other reports have shown that SOX2, in conjunction with OCT4, can directly regulate the expression of the miR-302 cluster in hESC, which can regulate the cell cycle.[266] SOX2 can directly regulate the expression of miR-137 during adult neurogenesis.[465] In a very recent study, SOX2 was silenced in glioblastoma multiforme (GBM), which identified 90 miRNA directly or indirectly regulated by SOX2.[466] This study also revealed the existence of a regulatory double-negative feed-back loop between miR-145 and SOX2.

The operation of OCT4 and NANOG and their transcriptional networks have been previously characterised, to varied extent, in 2102Ep and NTera-2 cells, but also in other ECC lines.[280,378,467–470] On the other hand, the SOX2 transcriptional network has only been characterised in the ECC line, NCCIT. The results in Chapter 4 showed that SOX2 downregulation could result in the differentiation of other types of ECC. The documented or potential importance of SOX2 in pluripotency, stem cell maintenance, CSCs and cancer led to the decision to profile its function in 2102Ep and NTera-2 cells. For this the gene and miRNA networks regulated by SOX2 in embryonal carcinoma were profiled by an integrated analysis of whole-genome and miRNAome expression in the two ECC lines after induced SOX2 downregulation. An additional aim of this project was to explore the possibility of correlating miRNA expression to mRNA expression, as this could lead to a large number of miRNA:mRNA relationships with potential importance to CSC, ECC and hESC function. MiRNA:mRNA cross-referencing has been used by different groups with varying methods. To apply an approach to a miRNA:mRNA correlation study, a number of important caveats have to be taken into account. Some of these caveats are the high number of false positive results produced by miRNA target prediction software and the influence of uncustomary miRNA function, e.g. coding sequence-binding or transcriptional activation. Another caveat is the often irregular inverse differential relationship between mRNA and miRNA (i.e. the inverted difference in fold-expression between miRNA and mRNA is often uncorrelated). Another problematic interference in the inverse correlation study of miRNAs and their targets is the previously documented observation that the epigenetic and transcriptional regulation of the expression of miRNAs and their targets is strongly linked, resulting in similar transcriptional expression of miRNAs and their functional targets.[471–473] Some of these difficulties have been highlighted in previous papers.[474,475] These caveats limit this type of study to a few general approaches which can be taken: a miRNA-mRNA correlation over a time course or an miRNA-

weighted target approach. This study opted for the second approach as the first one would have been prohibitively expensive to perform.

5.2- Chapter Aims

The primary aim of this chapter was to map genes and miRNAs directly or indirectly regulated by SOX2 in ECCs. This was achieved by differential whole-genome and miRNAome expression analysis after the siRNA-induced silencing of SOX2 in 2102Ep and NTera-2 cells. The second aim was to compare whole-genome and miRNAome expression of 2102Ep and NTera-2 cells with silenced SOX2, to determine whether similar SOX2-linked regulatory mechanisms exist in these cells. The third aim was to identify inverse relationships between miRNAs and their targets by statistical and qualitative means.

5.3 — Materials and Methods

5.3.1 — SOX2 silencing in NTera-2 and 2102Ep Cells

For whole-genome and whole-miRNAome analysis in NTera-2 cells after SOX2 knock-down, the 3-day post-transfection RNA and protein samples from Section 4.4.3 were used. The 2102Ep RNA samples from Section 4.4.2 were considered too old to guarantee RNA integrity, so new samples were created. For this, 2102Ep cells below 50 passages were cultured on 6-well plates at a density of 18,000 cells/cm^2 24 hours prior to transfection to yield an approximate confluence of 40% at transfection. 2102Ep cells were transfected with Silencer™ Select siRNA targeting SOX2 *(s13295, Ambion, Applied Biosystems, Life Technologies)*. The sequence for this siRNA can be found in Table 4.6. The cells were transfected using the forward transfection method described in Section 2.2.2.5 with the conditions described in Table 5.1. This table also lists the siRNA concentration and transfection agent volume used for NTera-2 cells in Section 4.4.2. For transfections in 2102Ep cells, three replicates of siSOX2, ssNC#1, Vehicle Controls and Non-Transfected controls were prepared. Duplicate replicates were prepared to ensure the recovery of sufficient RNA and protein. The cells were transfected for 6 hours and subsequently incubated for three days in 3 ml hEC medium in humidified conditions at 37°C and 5% CO_2. The medium was refreshed every 24 hours.

Table 5.1—Transfection agents and volumes used for siSOX2 transfection in 2102Ep and NTera-2 cells.

Cell line	Final siRNA concentration	Transfection agent (vol./well)
2102Ep	10 nM	Lipofectamine™ 2000 (10.57 µl)
NTera-2	30 nM	Lipofectamine™ RNAiMAX (8.46 µl)

5.3.1.1 — RNA Purification and Protein Extraction

RNA purification was performed with the *miR*Vana PARIS kit *(AM1556, Ambion, Applied Biosystems, Life Technologies)* as described in Section 2.3.2.3. Cells were lysed directly on the plates and purification was performed as per manufacturer's. RNA was eluted in 100 µl elution solution and stored at -80°C until use.

Protein extraction was performed with the Cell Disruption Buffer spiked with protease and phosphatase inhibitor cocktails as described in Section 2.6.2.1. Protein samples were stored at -80°C until use.

5.3.1.2 — Assessment of RNA and Protein Quantity and Quality

RNA quantity and purity was assessed with the Nanodrop 2000 spectrophotometer as described in Section 2.4.2.1. To ensure high quality of microarray results, it was crucial to use RNA of high

integrity (i.e. with low degradation). This is because microarrays are more sensitive to RNA degradation than qRT-PCR. To measure the integrity of RNA, the 2100 Bioanalyser *(Agilent Technologies)* microfluidics station was used in combination with the RNA Nano 6000 kit *(5067-1511, Agilent Technologies)*. RNA integrity of the RNA samples extracted from siSOX2 and ssNC#1-transfected cells was measured using the protocol found in Section 2.4.2.2.

Protein concentration was measured with the BCA method as described in Section 2.6.2.2.

5.3.1.3 — Gene Expression Analysis

To determine SOX2 expression, qRT-PCR was performed with the TaqMan® PCR system. The protocol for cDNA synthesis and qRT-PCR can be found in Section 2.5.2.1. Relative quantification was performed by measuring SOX2 and GAPDH expression in all samples using the Prism® 7900HT Sequence Detection System *(Applied Biosystems, Life Technologies)* and primer/probe assays shown in Table 5.2. SOX2 expression was normalised to GAPDH endogenous control expression.

Table 5.2—TaqMan® primer/probe assays used for gene expression analysis of SOX2 and GAPDH.

Gene	Assay ID
SOX2	Hs00602736_s1
GAPDH	4331182

5.3.1.4 — Protein Expression Analysis

SOX2 protein expression was measured by the Western blot technique. For each replicate, approximately 30 µg (2102Ep) and 26 µg (NTera-2) protein mixed with 6X loading buffer was loaded per lane of a 12% polyacrylamide gel. After adding 10 µl Prosieve Protein Color Marker *(50550, Lonza Ltd)* to a single lane, SDS-PAGE was performed as described in Section 2.6.2.3 and the transfer to membrane was performed for 75 minutes as detailed in Section 2.6.2.4. The membranes were blocked in blocking buffer for one hour and subsequently probed overnight at 4°C with anti-SOX2 antibodies diluted in blocking buffer as described in Section 2.6.2.5. After washing the membranes were incubated in secondary antibodies diluted in blocking buffer for one hour at room temperature. The antibodies used for Western blotting are listed in Table 5.3. The blots were visualised with the ECL technique described in Section 2.6.2.7. After this, membranes were stripped for 15 minutes as described in Section 2.6.2.8, blocked and reprobed for GAPDH expression.

Table 5.3—Primary and secondary antibodies used for protein expression analysis of SOX2 and GAPDH.

Gene	Predicted MW	1st Antibody ID	Clone Type	Host	Dilution	2nd Antibody ID	Dilution
SOX2	34.3 kDa	ab75485	Poly	Mouse	1:500	sc-2314	1:2000
GAPDH	36.1 kDa	ab8245	Poly	Mouse	1:25,000	115-165-062	1:5000

5.3.2 — Whole Gene Expression Analysis with Gene Arrays

Microarray analysis was performed with the 2102Ep and NTera-2 RNA samples obtained in Section 5.3.1. The RNA from siSOX2 and ssNC#1 was used for microarray analysis and one replicate was hybridised per chip. The experiment was performed in biological triplicate, therefore, per cell line, a total of 6 Affymetrix® Gene ST 1.0 chips were used.

5.3.2.1 — Analysis of RNA Integrity

RNA quantity and purity was measured prior to TaqMan® PCR analysis (see Section 5.3.1.2). RNA integrity was measured with the 2100 Bioanalyzer *(Agilent Technologies)* and the RNA Nano 6000 kit *(5067-1511, Agilent Technologies)*. The protocol was performed as detailed in Section 2.4.2.2. RINs were recorded from the electrophoretic trace curve where possible to produce a measure with which RNA quality can be determined and compared. As the software is sometimes unable to deduce RINs, a visual inspection of the electrophoretic trace curve was performed to determine RNA quality.

5.3.2.2 — Sample Preparation, Hybridisation and Scanning

Sample preparation with the WT Expression Kit *(4411973, Ambion, Applied Biosystems, Life Technologies)* was performed in parallel for all six samples per cell line as described in Section 2.5.2.5. In a single experiment, Control RNA was used. Control RNA, which is included in the GeneChip® WT Terminal Labelling and Controls Kit *(901525, Affymetrix®)*, can, when used in the case of one experimental RNA sample, provide a comparative standard to assess successful sample preparation. Chip hybridisation and scanning was performed for all six samples in parallel. Washing the chips was performed with three samples at a time with, with the three replicates of each sample type washed in parallel. Scanning was performed in similar a manner.

5.3.2.3 — Quality Control

Quality control of the Affymetrix® Gene ST 1.0 microarrays was performed after scanning of the chips and recording of the data. These controls were performed to assess the integrity of the sample RNA and the precision of the sample preparation and array hybridisation steps. Two stages of quality control were performed: visual inspection of the chip image and computational inspection of certain negative and positive controls located on the Affymetrix chip. Visual inspection was performed with the Viewer Module of the GeneChip® Command Console software *(Affymetrix®)*. Computational quality control was performed with the Expression Console™ software *(Affymetrix®)*. For quality control analysis Expression Console™ applies the summarisation method, Robust Multi-Array Average (RMA) analysis, to summarise probe

intensities into probe sets. For descriptions of both stages of quality control see Table 5.4 and Table 5.5. Sections 2.5.1.3 and 2.5.2.5 give a detailed background of aspects of the *Affymetrix® GeneChip®*.

Table 5.4—Visual quality control of the gene arrays.

Visual Inspection	
Whole-chip inspection	The chip is inspected for scratches, speckles and lightened/darkened areas due to unequal hybridisation or air bubbles.
Oligo B2 control inspection	The oligo B2 control is used by the scanning software as waypoints from which the topography of the chip can be analysed. Inspection of the oligo B2 hybridisation should be visible as a fluorescing checkerboard corner and alternating borders along the array. This is crucial for a correct grid alignment of the individual probes. As an additional control for correct olgo B2 hybridisation, the name of the chip should show in its complete entirety, with no missing gaps.

Table 5.5—Computational quality control of the gene arrays.

Computational Inspection	
Probe Level Metrics	This quality assessment metric is conducted with the probe level data: - **pm_mean** is the raw intensity mean for all PM probes on the array before any data transformations or adjustments have been performed on the intensity data. With this outliers with unusually bright or dim overall intensities can be established from the overall sample set. These chips are generally not an issue, but they can be flagged for closer inspection during further quality control analysis. - **bgrd_mean** is similar to pm_mean, but instead applies to the mean of raw intensity of all background probes. The overall intensity of these probes are correlated to those of the PM probes and might not necessarily be lower than the pm_mean, as the bgrd_mean metric could be skewed due to high GC count background probe bins. However, high bgrd_mean with low pm_mean can suggest poor sample data, especially if uncorrelated among samples.
Probeset Summarisation Based Metrics	- **pos_vs_neg_auc** is the area under the curve for a receiver operating characteristic plot, which compares signal values for the positive controls to the negative controls. The probe signals of the positive controls, which ideally are exclusively true positives, are expected to be separated from those of the negative controls, which ideally are false negatives. A pos_vs_neg_auc value of 1 indicates perfect separation, whereas 0.5 would suggest no separation. Values between 0.8 and 0.9 are typical. - **X_mean** is the signal mean for the chosen category of negative or positive controls (X). - **X_mad_residual_mean** is the mean of absolute deviation of the residuals from the median i.e. the extent to which probes hybridising to a common target behave differently from the norm predicted by the robust RMA algorithm (see Section 5.3.2.4). The residual is the difference between the actual and predicted value. Every probe will have a typical residual within a robust probe model and if this residual is very different from the median it could indicate to a problematic chip. X is the category of controls.

—Table continued—	
	- **X_rle_mean** is the relative log expression for all probe sets within category X. This compares the estimated signal of a probe set on a given chip to the median signal of this probe set over all chips and log_2-transforms the difference. The relative log expression is calculated for all probe sets in category X and when analysed together should be consistently low among similar samples, reflecting low biological variance between replicates.
Probe set list	- Checking individual raw probe set values can indicate expected behaviour. In the case of this study consistent expression among endogenous controls such as GAPDH and B2M is expected and probe set values of SOX2 and also OCT4 and NANOG are expected to be lower in the siSOX2 transfected samples. - In addition to the individual probe sets, collective probe set sub-groups are also checked. These are: o **all_probeset**, a sub-group which comprises the mean value of all probe sets on the chip. o **bac_spike**, the pre-labelled bacterial spike control probe sets, which can indicate hybridisation status. o **polya_spike**, the set of polyadenylated RNA spikes, which can indicate the status of the target preparation. More variability in this sub-group is expected due to a limited number of spike controls and probe sets. o **neg_control**, a sub-group which target introns of endogenous control genes. This large group of probe sets should have a low signal, unless significant intron RNA is retained in the samples. o **pos_control**, the sub-group of probe sets targeting exonic regions of genes, which has a value depending on the amount of genes transcribed.
Pearson's Correlation	A graphical display of the correlation of raw intensity values of all probes between all chips. Correlation among replicates of negative or positive control samples is expected to be higher than the correlation between the opposite controls.
Hierarchical clustering	A graphical display of results more correlated replicates clustered in a hierarchy based on similarity. Each branch-end represents the background-adjusted and normalised results from a single chip.

5.3.2.4 — Probe Signal Summarisation, Background Correction and Normalisation.

With RMA analysis individual raw probe intensity values were summarised, corrected to background signals and subsequently normalised to internal controls. RMA analsysis is widely used mathematical and statistical formula for the analysis of microarrays from various different manufacturers, including Affymetrix®. It consists of several steps described in the paragraphs below. Other algorithms exist, such as older chip-by-chip methods such as MAS5 to newer project-based model-fitting methods such as dCHIP. At the time of this study no published, extensive comparison of different expression algorithms for the Affymetrix GeneChip® Gene 1.0 ST could be found. However the GeneChip® Exon 1.0 ST arrays are quite comparable to this chip due to similar underlying principles of design. A key difference between these chips is that the majority of poor-performing probes found on the GeneChip® Exon 1.0 ST array is absent on the GeneChip® Gene 1.0 ST array. Still, the high operational and experimental similarities between these array types were previously validated and this was taken into account when selecting a method of analysis.[300] RMA analysis has a number of benefits over other methods, but also a few caveats. A previous study described RMA analysis to have greater precision than most other methods, but at a cost of underestimating fold change.[476] It increases sensitivity to detection of small expression changes, however it has an increased chance of producing false positives compared to hybrid algorithms such as PLIER for poorly performing probes. It was reasoned that as the GeneChip® Gene 1.0 ST array contains probe sets which were selected by the manufacturer to be high performing, that the latter issue would be less relevant, thus providing the added benefits of RMA analysis while reducing the burdens.

For this chapter, standard RMA analysis was performed computationally by colleague, Dr. Gordon Blackshields with the open-source statistical analysis tools, Bioconductor 2.9 and CRAN, also called 'R'. The first step of the RMA method is the array normalisation of probe values across the replicates using full quantile normalisation as described in previous research.[477] After this the intensity values produced by the remaining probes were ln(x) transformed. Subsequently, background correction was applied to reduce noise. This was performed by subtracting from all probes, the median expression score of each strata of background probes with comparable CG count. Finally the individual probes are summarised into probe sets by the application of the Tukey's median polish to all input hybridisations which were expressed above the background expression, as determined by a one-sided t-test.

5.3.2.5 — Computational Analsysis of Differential Gene Expression

For the identification of differentially expressed genes the Bioconductor open-source software package, limma (Linear Models for Microarray Data), was used.[478] This free and open software package is widely used and highly curated for the analysis of many types of microarrays from different manufacturers. RMA-transformed gene array data produced previously was interpreted by limma using three distinct functions:

1. The **lmFit** function fitted a general linear model for each gene to calculate its fold change and standard error.
2. The **eBayes** function then applied the parametric empirical Bayes method to test for transcripts with log-transformed fold changes that exceed a threshold.
3. The **topTable** function then presents a list of genes that have the highest lilkelyhood of differential expression. These are sorted and filtered by a P-value, which is adjusted for multiple testing with the Benjamini-Hochberg false discovery rate (FDR).

For this study, any genes which were significantly >2-fold differentially expressed with a maximum adjuste P-value of 0.05 were considered deregulated and included in further analysis.

5.3.2.6– Computational Analysis of Gene Ontology

In this chapter gene ontological analysis was performed on the differential gene expression and the predicted targets of differentially expressed miRNA. Ontological gene analysis is a bioinformatic method to assign particular attributes and annotations to genes which are descriptive of the biological function and properties of these genes and to do this in a code that is universally understood. These universal gene annotations are multi-layered according to domain and functional hierarchy. As more functions about a particular gene are discovered, it can follow the hierarchy to be given more specialised annotation. Ontological gene analysis was initiated by the Gene Ontology (GO) project, which assigns GO-terms to a very large number of genes and gene products in many different organisms. Since this initiative, many similar biological databases have emerged with broadly similar tenets, such as KEGG and PANTHER. As the annotations are universally adaptable, many different bioinformatic tools utilise them for high-volume analysis of gene lists. For gene annotation in this study, DAVID Bioinformatic Resources 6.7 was used. This is a tool developed by Laboratory for Immunopathogenesis and Bioinformatics for the National Institute of Allergy and Infectious Diseases which can statistically detect an enrichment of functional annotations in gene data. DAVID is one of the most widely used tools for the identification of shared functional annotations and pathway relationships in high through-put gene data.

Functional annotation analysis of genes was performed with the aid of general guidelines outlined by the developers of the software package.[479] This tool was used for this study as it can call on a large amount of annotation categories, which, besides traditional GO terms, also contain categories from other databases, such as KEGG and BIOCARTA. From large gene lists, DAVID can produce functional annotation tables, which are ranked according to significance as determined by the EASE score. This score is calculated from a Fisher's Exact P-value, which is modified to provide a more conservative value. For this calculation, the number user-included genes belonging to a similar pathway/annotation, as a proportion of all genes in this pathway/annotation is tested against all user-included genes that are not part of this pathway/annotation as a proportion to all the known genes in the genome. In addition to the EASE score, a value is also given based on the multiple test correction of the EASE score as calculated by Benjamini-Hochberg FDR. These values can indicate and statistically quantify the enrichment of particular annotations in the entire dataset.

For functional annotation analysis certain parameters were set in DAVID. A particular annotation was considered enriched when the EASE-score (P-value) ≤0.05 for most annotations within annotation clusters. For multiple comparison correction, Benjamini-Hochberg FDR was considered during analysis, but was not critical for determining the importance of certain annotations in light of their expected biological contribution. For some gene lists, functional annotation clustering was used with 'Classification Stringency' set at 'Medium'.

Due to the various different annotation systems used and the similarity between various annotations of the same system, the functional annotation results produced by DAVID can be highly redundant (i.e. many annotations overlap in description). To reduce this redundancy and focus on important cellular functions and pathways, DAVID provides functional annotation clustering, based on statistically measuring the degree of common genes between two annotations and a subsequent heuristically grouping of similar annotations in to clusters. The statistical enrichment of these annotations is assessed with an EASE score as described in the previous paragraph. In addition, the enrichment of clusters is assessed by giving each cluster a Group Enrichment Score, which is the –log geometric mean of all members' EASE scores. A Group Enrichment Score of 1.3 or higher represents a P-value of 0.05 and lower. The (corrected) EASE scores for individual annotations and clusters are not the sole method of interpreting the data, as biological importance of particular functions and pathways have to be taken into account. For example, an annotation or cluster with a relatively low significant enrichment could still be considered to the experimented biology.

In this chapter, the functional annotation tables are shown for gene lists which show a high amount of significant enrichment per individual annotation. For gene lists which show little or no significant annotation, functional annotation clustering is adopted. This approach was used to include all significant functional annotations and avoiding the unintended omission of important annotations which could appear in low-ranking clusters. Functional annotation clustering was considered very useful to find significant enrichment in more broadly annotated data with low scoring individual annotations as it could emphasise the biological interpretation to important annotations by adding an additional layer of functional grouping. In other words, a very specialised approach, producing a collection of statistically insignificant results, was superseded by an approach that could find a more general, but also more biologically relevant enrichment. Additionally, because functional annotation can produce very long lists of results, the use of functional annotation tables for highly significant enrichments was preferred over clustering as a more efficient method of displaying relevant information in the finite constrictions of this document.

5.3.3- TaqMan® Array Human MicroRNA

The RNA from the SOX2 silencing experiment in 2102Ep and NTera-2 cells, described in Section 5.3.1, was used for whole-miRNAome expression analysis with the TaqMan® Array Human microRNA A+B Cards Set v3.0 *(4444913, Applied Biosystems, Life Technologies)*.

5.3.3.1 — Optimisation of the Preamplification Stage

Prior to use, the unamplified and pre-amplified miRNA batches were tested for miR-16 expression to assess whether preamplification was necessary. Section 5.4.7.1 briefly describes the experiment.

For this experiment miRNA cDNA was synthesised from 210 ng total RNA from each sample using the MegaPlex™ Primer Pools A and the manufacturer's protocol. Subsequently, a part of this cDNA was preamplified with MegaPlex™ Preamp Primers A. The preamplified cDNA was diluted 1:4 in 0.1X TE buffer as per manual. Using a conventional TaqMan® miRNA Assay for miR-16, unamplified and preamplified cDNA was assayed for miR-16 expression with the method described in Section 2.5.2.3. One technical replicate per biological replicate was used.

5.3.3.2 — Whole miRNA Expression Analysis

For whole-miRNAome expression analysis the purified and quantified total RNA from siSOX2 and ssNC#1-transfected 2102Ep and NTera-2 cells was used. This was from the same batch as used for the whole-genome analysis experiment. Preamplification and cDNA synthesis was conducted with the Megaplex™ PreAmp Primers, Human Pool v3.0 and Megaplex™ Primer Pools, Human Pools Set v3.0 respectively as described in Section 2.5.2.4. From 2102Ep 105 ng and from NTera-2 cells 210 ng total RNA was used for each 7 µl cDNA reaction. These RNA quantities were determined to produce optimal results from each cell line, as is explained in Section 5.4.7.1. For each sample two TaqMan® Array cards were used, one A and one B card, each containing wells with unique lyophilised TaqMan® miRNA primer/probes. Effectively, one technical replicate was used per biological replicate. Three biological replicates were run per sample. QRT-PCR was performed with the TaqMan® Arrays as per Section 2.5.2.4.

5.3.3.3 — Computational Analysis of Differential miRNA Expression

The raw data from the qRT-PCR runs were collected and imported in to the RQ Manager ver. 1.2.1 software *(Applied Biosystems)*. This tool was used to collectively appoint threshold and baseline values to all amplification plots simultaneously across all replicates tested. Approximately half of the threshold and baseline values were successfully automatically established by the software, while manual adjustments were needed for the rest. In addition to this, unexpressed miRNAs or miRNAs with highly aberrant amplification plots were omitted. The exported results were subsequently imported into DataAssist ver. 3.0 *(Applied Biosystems)*, a software tool designed to perform relative quantification across multiple assays with the $2^{-\Delta\Delta Ct}$ method. The raw Ct values exported from the RQ Manager software was imported into DataAssist. From this the six samples (three experimental and three negative control replicates) were filtered by the software to bring any raw expression values with Ct > 40 down to 40 Ct and include these in the calculations. The software provides a modified Grubb's test to remove outliers from technical replicates. However, this was only applied, when necessary, to the U6 small nuclear RNA results as these are the only results with multiple technical replicates on each card. Two endogenous controls were selected for data normalisation. These were U6 snRNA and miR-16. U6 snRNA was the standard endogenous control represented on both A and B cards and its expression is measured by multiple technical replicates on each card. This snRNA is selected by the manufacturer due to its similarity to miRNA in size and due to its common and stable expression in many different cell and tissue types. MiR-16 was selected as an additional endogenous control on the A cards due to its previous successful use in 2102Ep and NTera-2 experiments during this study and others. As miR-16 is not represented on the B cards, miR-16-1* was used as an endogenous control in NTera-2 cells. This miRNA is the antisense minor sequence of miR-16, produced from the same stem-loop, pre-miR-16. MiR-16-1* had similar expression to miR-16 on the A cards and U6 snRNA on the B cards. The software calculated the mean Ct value of U6 snRNA and miR-16/miR-16-1* expression to which the data were normalised to produce a ΔCt for each miRNA. Interestingly, miR-16-1* was not expressed in 2102Ep cells. This is not uncommon, as minor sequence miRNAs are often rapidly degraded. For normalisation of 2102Ep B card data, global normalisation was chosen, as this was considered to be more reliable and more accurate than U6 snRNA expression alone in this case. The $2^{-\Delta Ctexp.}$ was calculated from each ΔCt value of the experimental control set and the $2^{-\Delta Ctneg}$ was calculated from each negative control ΔCt. To calculate the expression fold-difference, the geometric mean of the $2^{-\Delta Ctneg}$ for each biological replicate of each miRNA was subtracted from the corresponding individual $2^{-\Delta Ctexp}$. A P-value was calculated with a two-tailed Student's t-test comparing the two groups. A respective signal correlation was created by calculating a Pearson's product moment correlation coefficient for the Ct values of sample pairs.

5.3.3.4 — Computational miRNA Target Prediction

This was performed by TargetScan v5.1 as detailed in Section 3.3.9.

5.4 — Results

5.4.4 — Silencing of SOX2 in 2102Ep and NTera-2 Cells

SOX2 was silenced in both cell lines with the method described in Section 5.3.1. Transfection was conducted with the forward transfection method of SOX2-targeting Silencer™ Select siRNA and ssNC#1 in both 2102Ep and NTera-2 cells. Vehicle and non-transfected controls were included. The cells were lysed three days post-transfection and total RNA and protein was obtained.

After RNA purification, quantity and purity were assessed with the Nanodrop 2000 spectrophotometer as described in Section 2.4.1.1. Gene expression was measured with the TaqMan® system as described in Section 5.3.1.3. SOX2 expression was measured with relative quantification to the endogenous control, GAPDH.

Protein expression was performed with the Western blot technique as described in Section 5.3.1.4. All samples were probed for SOX2 and GAPDH expression.

5.4.4.1 — SOX2 Expression in 2102Ep

SOX2 gene expression results are displayed in Figure 5.1. SOX2 silencing was successful with only 7% remaining SOX2 expression in 2102Ep cells transfected with 10 nM siSOX2. No significant difference existed between the ssNC#1-transfected cells and the non-transfected controls, however SOX2 expression in the vehicle controls was significantly higher than in ssNC#1-transfected cells (P = 0.0356). As this difference was small (33%) and as no significant difference existed between vehicle controls and non-transfected controls, this was not considered important. Protein expression, as shown in Figure 5.2, confirmed the silencing of SOX2 in siSOX2 transfected 2102Ep cells.

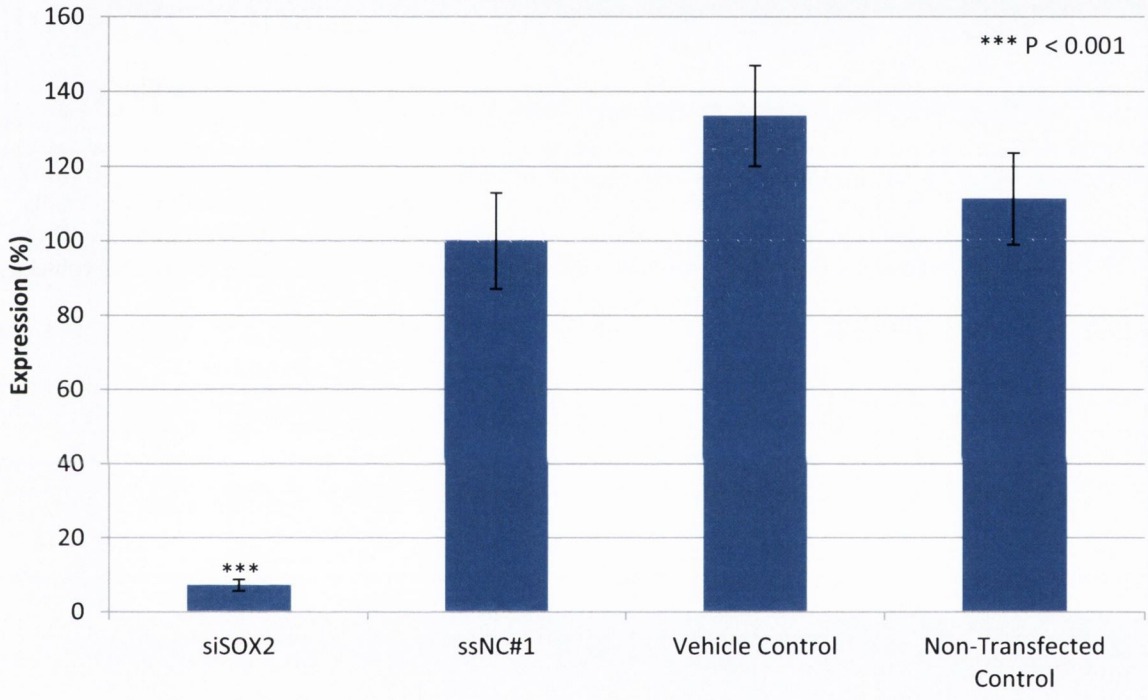


Figure 5.1—SOX2 expression in 2102Ep cells three days after transfection with 15 nM siSOX2. SOX2 expression was significantly lower in siSOX2-transfected cells compared with ssNC#1-transfected cells (P = 0.0002). The SOX2 expression in the latter samples was set at 100%. All experiments were performed in biological triplicate and the values shown is mean relative expression of these triplicates ± SD. A t-test was used to compare sample means.

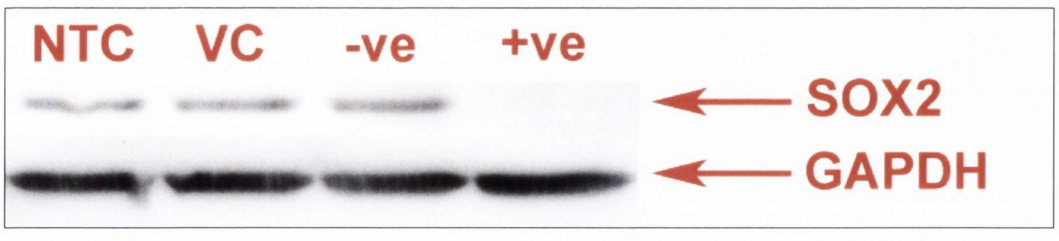


Figure 5.2—Western blot analysis of SOX2 expression in siSOX2 transfected 2102Ep cells. From right to left: siSOX2, non-targeting, vehicle and non-transfected controls. +ve are the siSOX2-transfected cells. –ve represents the ssNC#1-transfected cells. VC and NTC are the vehicle and non-transfected control respectively.

5.4.4.2 — SOX2 Expression in NTera-2 Cells

For NTera-2 cells, the sample RNA and protein from Section 4.4.2 were used. At the mRNA level, 49% of remaining SOX2 expression was detected. To assess whether SOX2 protein expression was similarly high or lower, Western blots were performed. The results showed a clear silencing of SOX2 in the siSOX2 transfected cells (see Figure 5.3). This was fully confirmed by further increasing the exposure time of the blot, which showed no SOX2 bands in the siSOX2 transfected samples. This confirmed that SOX2 was silenced in NTera-2 cells three days after siSOX2 transfection.

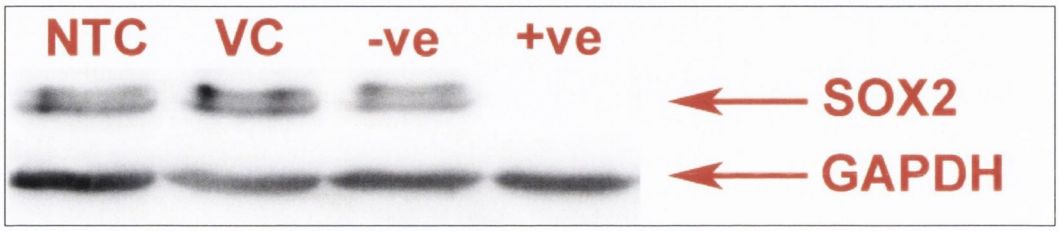


Figure 5.3—Western blot analysis of SOX2 expression in siSOX2 transfected NTera-2 cells. From right to left: siSOX2, ssNC#1, vehicle and non-transfected controls. +ve are the siSOX2-transfected cells. –ve represents the ssNC#1-transfected cells. VC and NTC are the vehicle and non-transfected controls respectively.

5.4.5- RNA Quality and Quantity Assessment

The original spectrophotometric A_{260}/A_{280} and A_{260}/A_{230} ratios, obtained earlier, are displayed in Table 5.6 and Table 5.7. A_{260}/A_{280} and A_{260}/A_{230} ratios between 1.9 and 2.1 are considered pure i.e. free from significant protein and organic solvent contamination respectively. As the DNA digestion during RNA purification decreases these ratios slightly, the ratios displayed in the tables are on average 0.1-0.2 points higher previous to DNase treatment. According to the manual of the Turbo™ DNA-*free* kit used for DNA digestion, this should not affect downstream experiments such as microarray analysis. A_{260}/A_{280} in all cell lines in both samples were within 1.9-2.1 range.

To ensure only high quality RNA was used for microarray analysis RNA integrity was measured with the 2100 Bioanalyzer *(Agilent Technologies)* microfluidics station. Visual inspections of the electropherograms produced from 2102Ep and NTera-2 RNA by the 2100 Bioanalyzer showed no presence of RNA degradation products, indicating that RNA integrity was sufficient (see Figure 5.4 for representative samples). As a comparison, a figure provided by Agilent was included showing electropherograms from RNA samples with various degrees of integrity (Figure 5.5). This clearly showed that the electropherograms from the 2102Ep and NTera-2 RNA samples matched the highest quality example. As a secondary validation of RNA quality, the automatically calculated RINs are included in Table 5.6 and Table 5.7.

RINs for 2102Ep samples were within a 9.7-10 range, indicating very high RNA integrity. For the NTera-2 samples, the software could not generate a RIN (even after adjusting threshold properties) as an analysis error indicated an anomaly in the electropherogram. It is possible that this was caused by a low overall fluorescence. However, the visual inspection of electropherograms produced for 2102Ep and NTera-2 RNA indicated that the sample quality was sufficient for gene array analysis.

Table 5.6—RNA purity and integrity in 2102Ep samples.

Sample	A_{260}/A_{280}	A_{260}/A_{230}	RIN
siSOX2 Rep. 1	1.98	1.81	10
siSOX2 Rep .2	1.96	1.92	10
siSOX2 Rep. 3	1.93	1.95	10
ssNC#1 Rep. 1	1.97	2.03	10
ssNC#1 Rep. 2	1.96	2.04	9.7
ssNC#1 Rep. 3	1.95	1.93	10

Table 5.7—RNA purity and integrity in NTera-2 samples.

Sample	A_{260}/A_{280}	A_{260}/A_{230}	RIN
siSOX2 Rep. 1	1.99	1.95	n/a
siSOX2 Rep .2	2	1.75	n/a
siSOX2 Rep. 3	2.01	1.91	n/a
ssNC#1 Rep. 1	2.01	1.88	n/a
ssNC#1 Rep. 2	1.99	1.92	n/a
ssNC#1 Rep. 3	1.98	1.89	n/a

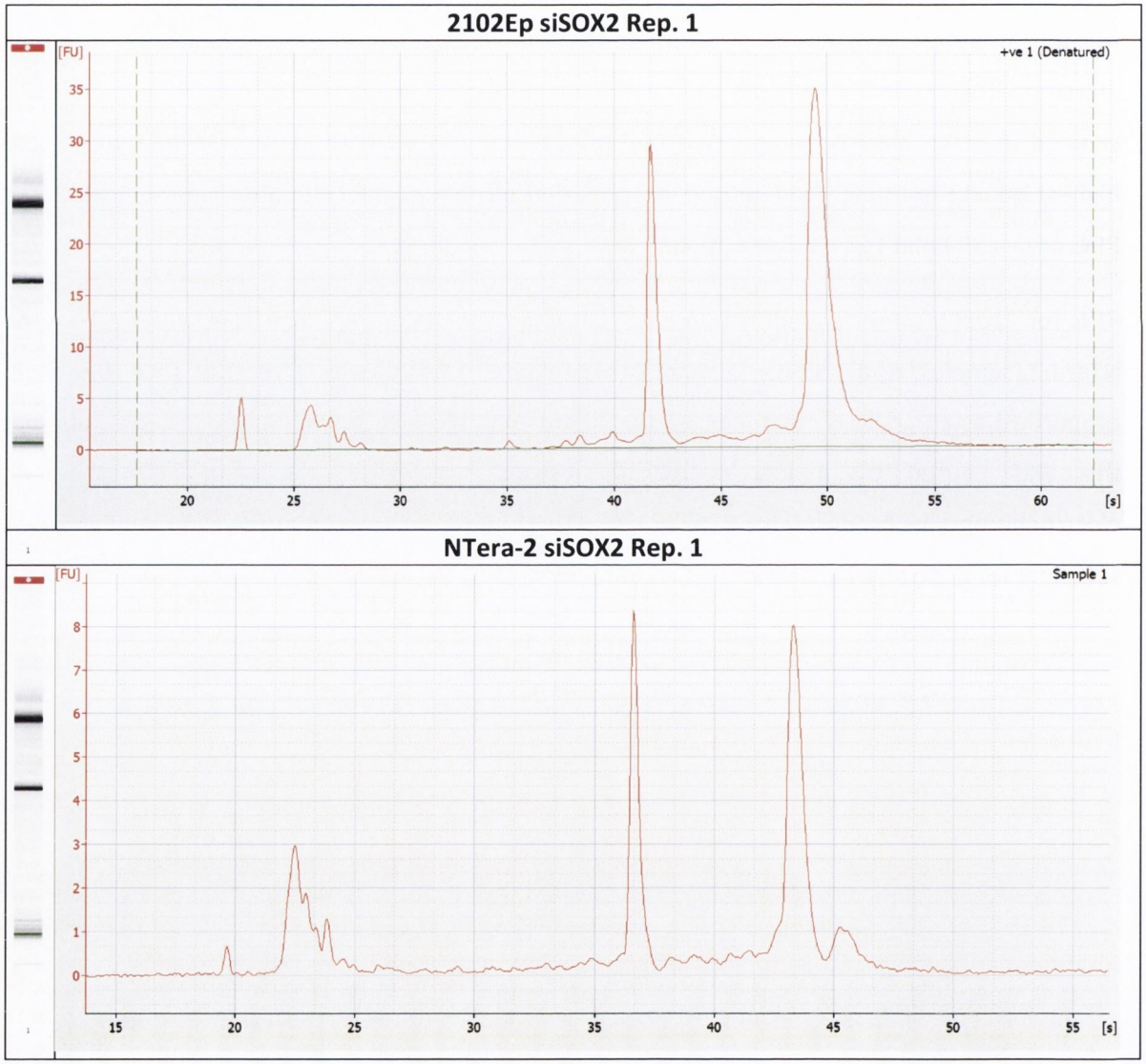


Figure 5.4—Electropherograms of single 2102Ep and NTera-2 RNA samples. The ladder on the left is the electrophoretic trace. The peaks in the graphs represent the fluorescent bands in the ladder. The figures shown here were representative of the entire sample set in each cell line.

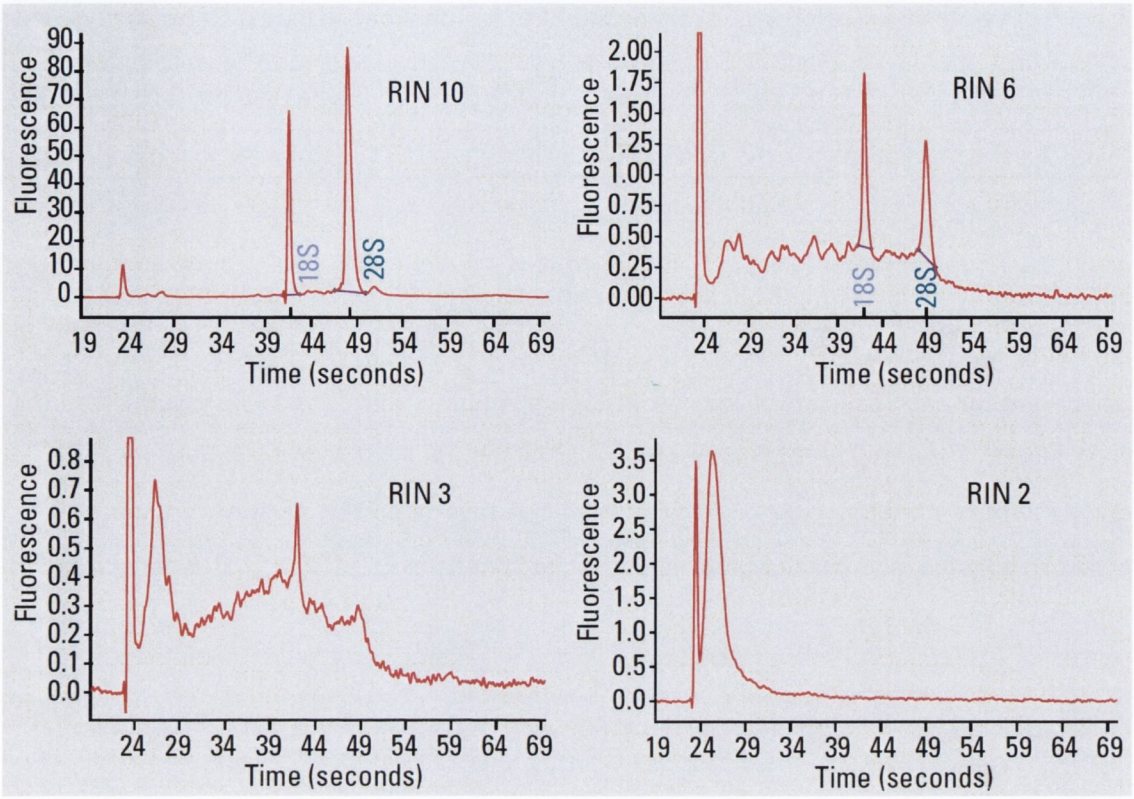


Figure 5.5—Example of electropherograms from RNA with different integrities. RNA integrity decreases with decreasing RIN values. RIN 10 indicates highly intact RNA. RIN 2 is representative of highly degraded RNA. *(Image source: Agilent Technologies)*

5.4.6 — Whole-Transcriptome Differential Expression Analysis with Gene Arrays

This Section contains the results from the total differential gene expression analysis in 2102Ep and NTera-2 cells, three days after silencing SOX2 in both these cell lines.

5.4.6.1 — Quality Control

For both 2102Ep and NTera-2 cell lines, RNA sample preparation, chip hybridisation, washing and scanning was performed with RNA from siSOX2 and ssNC#1-transfected cells. This procedure was performed for each cell line sample set at different times, with a total of 6 samples during one experiment. The protocol for this can be found in Section 5.3.2.2. After data collection, a set of quality control analyses were performed with the raw data files. Quality control analysis was performed by the methods described in Section 5.3.2.3.

Except for one sample (NTera-2 siSOX2 replicate 3), the initial visual inspection of the chip scans, shown in Figure 5.6A, showed no anomalies or lightened or darkened areas on the chips due to unequal hybridisation. Chip hybridisation and scanning was repeated for the faulty sample, after which a clean visual result was obtained. In addition to an overall inspection, specific regions of the chip were checked for oligo B2 control hybridisation (see Figure 5.6B). The purpose of this control is primarily for grid alignment during the measurement of probe fluorescence. It also serves as a control for correct hybridisation. The corner regions of the chip show a typical checkerboard pattern and the sides of the chip show an alternating pattern. The bottom-left region shows the complete name of the chip (see Figure 5.6C. Lastly, the alignment grid perfectly matches the position of the oligo B2 control probes. The visual inspection of the gene array chips preliminarily show correct sample hybridisation and grid alignment.

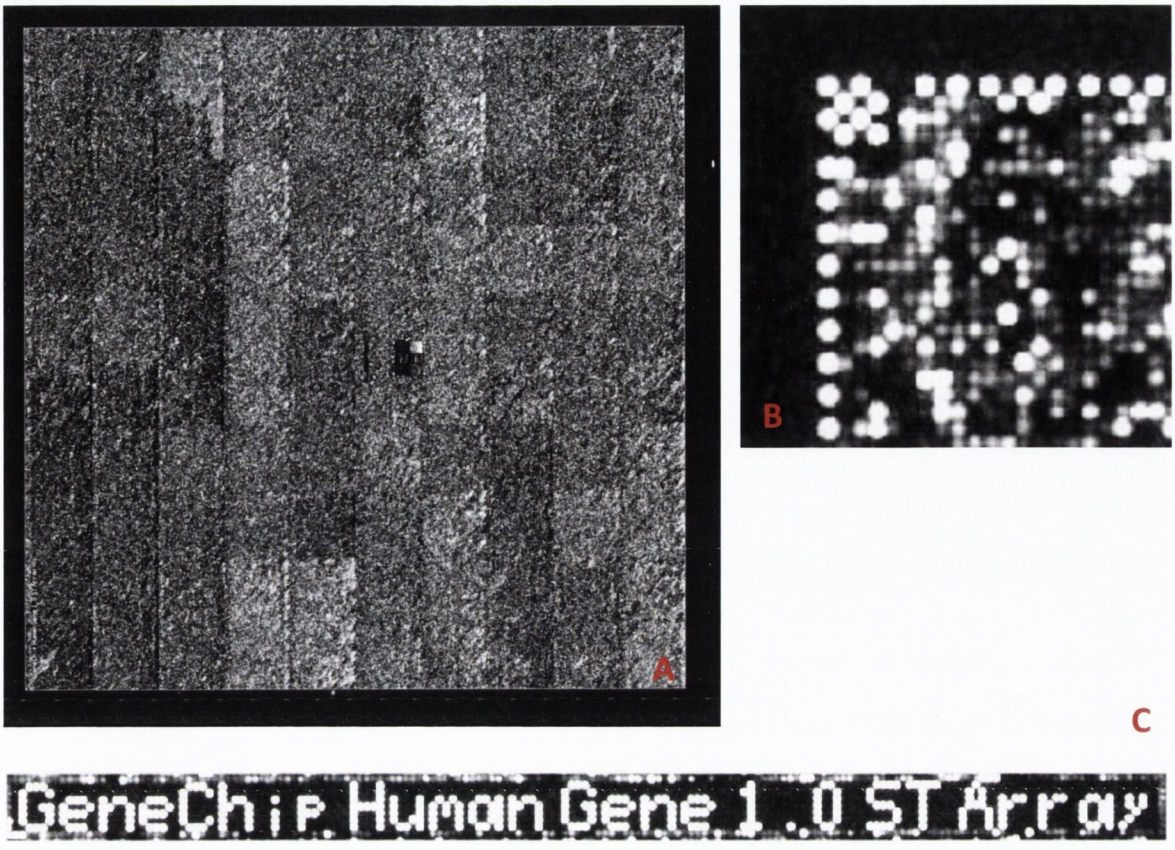
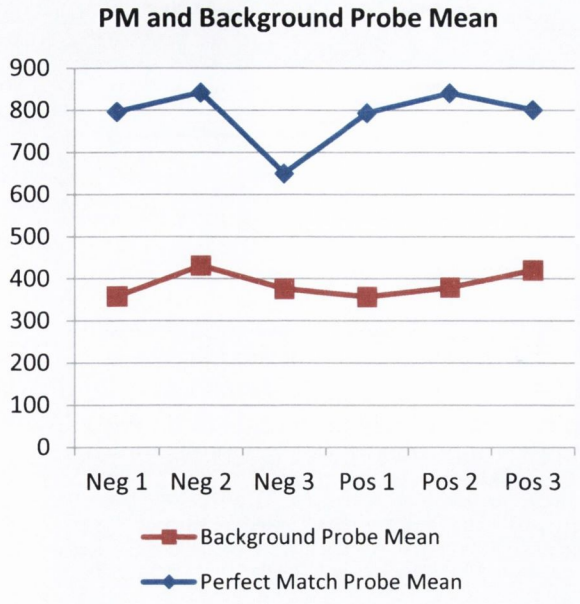


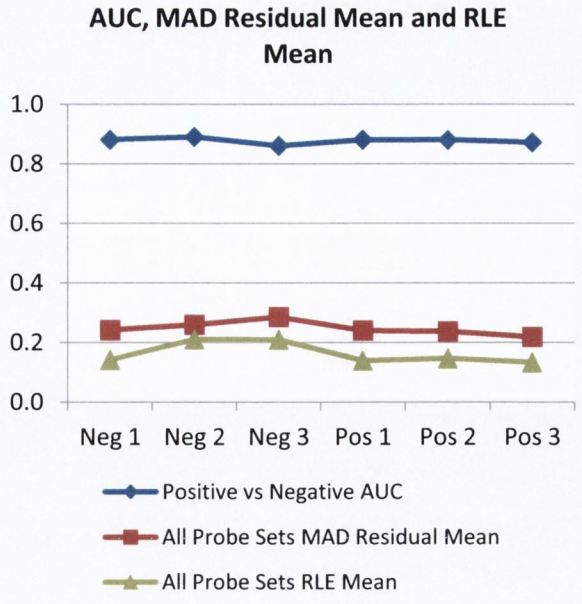
Figure 5.6—Visual quality control of the GeneChip® Human Gene 1.0 ST array. A picture of the whole array indicates the abscence of bubbles, specks or other visual anomalies (A). A checkerboard pattern in the corners of the chip and dotted boarders (B) as well as the full appearance of the array's name (C) indicate correct hybridisation of the oligo B2 control probes.

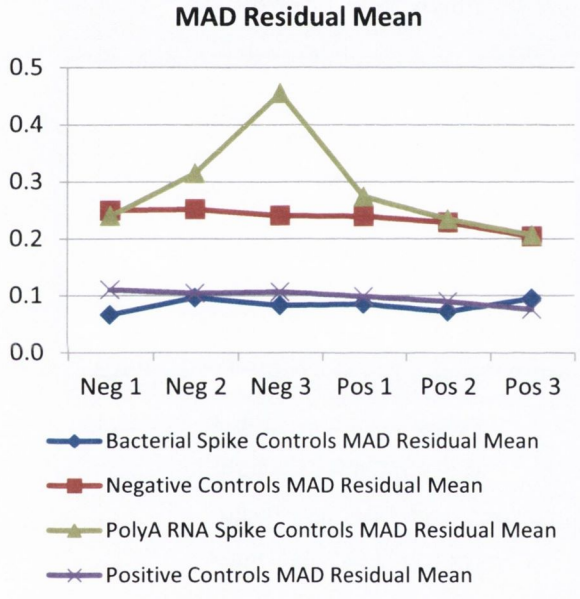
The next stage consisted of probe level and probe set metric analysis to inspect the fluorescence of certain probe and probe set subcategories prior to background adjustment and signal normalisation. A description of quality control metrics used is described in Section 5.3.2.3. The various quality control metrics used are displayed in Figure 5.7 and Figure 5.8.

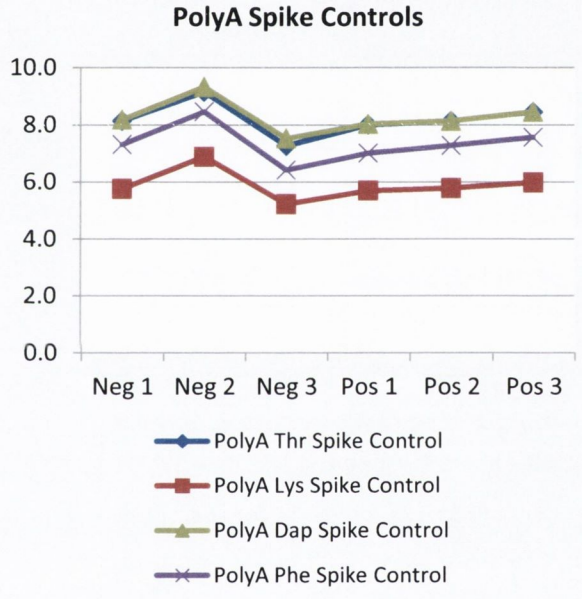
The probe level metrics measured were the perfect match (targeting probes) and background probe mean (see 'PM and Background Probe Mean'). Overall chip intensity (perfect match probe intensity) was amply sufficient for both 2102Ep and NTera-2 cells. Replicates 2102Ep Neg 3 and NTera-2 Pos 3 were slightly less fluorescent than the other replicates, but within acceptable limits. Variance among background probe fluorescence was quite low in all replicates and intensity was significantly less than that of the perfect-match probes. For both perfect-match and background probes, any variability was compensated by the RMA analysis. Next, metrics were analysed based on probe set categories (see 'AUC, MAD Residual Mean and RLE Mean'). The AUC of the positive controls compared to the negative controls was high for both cell lines, with a mean of 0.88. This reflects sufficient separation between positive and negative controls as values between 0.8 and 0.9 are typical. The MAD residual mean (mean of absolute deviation of the residuals from the median) and relative $\log_2$ expression (RLE) mean of all probe sets were low as values between 0.1

and 0.4 are typical. The low MAD residual mean indicates that probes are in reality behaving very similarly to the model predicted by the RMA algorithm. The RLE mean is a similar metric, but displays the mean of the $\log_2$ deviation of the signal generated by each probe set across all chips from a predicted median value of said probe set. Low RLE mean reflects low biological variability among the replicates. As the cells were not given much time to undergo dramatic phenotypic change after siSOX2 transfection, a low RLE mean was expected and resulted. This result was also reflected in Figure 5.9, in which the width of the boxes is proportional to the RLE mean. Figure 5.9 also shows skewness as the median of the boxplot, which in all samples for both cell lines was very close to zero, indicating little skewness or "signal bias". When the overall MAD residual mean was deciphered to its component parts, a high value for the PolyA RNA Spike control was observed in 2102Ep Neg 3 (see 'MAD Residual Mean'). Because the corresponding metrics (the MAD residual mean of the Bacterial Spike, Positive and Negative controls) were all fine, this outlier was probably caused by a pipetting error during the addition of the PolyA RNA Spike control. The PolyA Spike and Bacterial Spike controls (see respective graphs in Figure 5.7 and Figure 5.8) identify any problems during the sample preparation and chip hybridisation stages of the protocol respectively. The PolyA Spike controls should be ranked *Dap>Thr>Phe>Lys*. This is the case indicating no problems during sample preparation, although in both cell lines *Dap* and *Thr* are quite proximate in both 2102Ep and NTera-2 arrays. This artefact could not be fully explained, but after consultation with Affymetrix, the manufacturer of the PolyA Spike controls, it was clarified that this was a known issue which could sometimes occurs with the WT Expression Kit. Affymetrix considered this inconsequential to the quality of the IVT stage or the overall protocol and they, furthermore, commented positively on the high quality of the other metrics used. Finally, the ranked order of the Bacterial Spike controls is as expected (CreX>BioDN>BioC>BioB) indicating that chip hybridisation and washing were performed correctly.









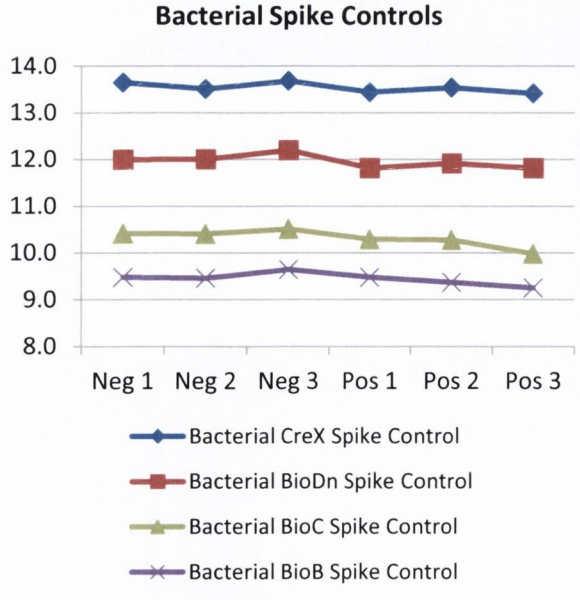
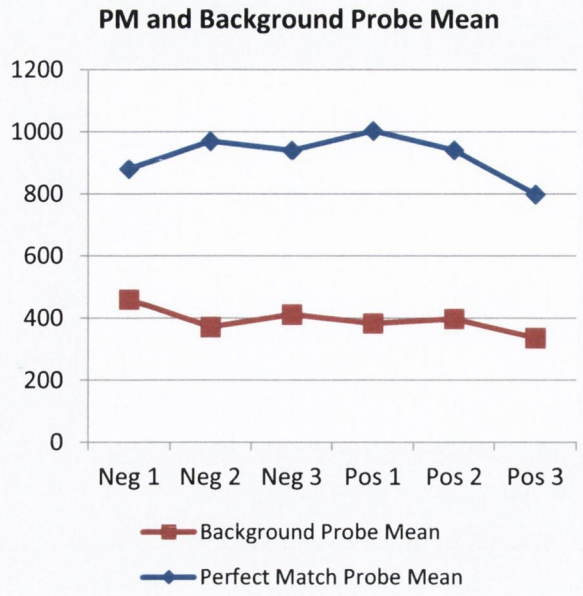
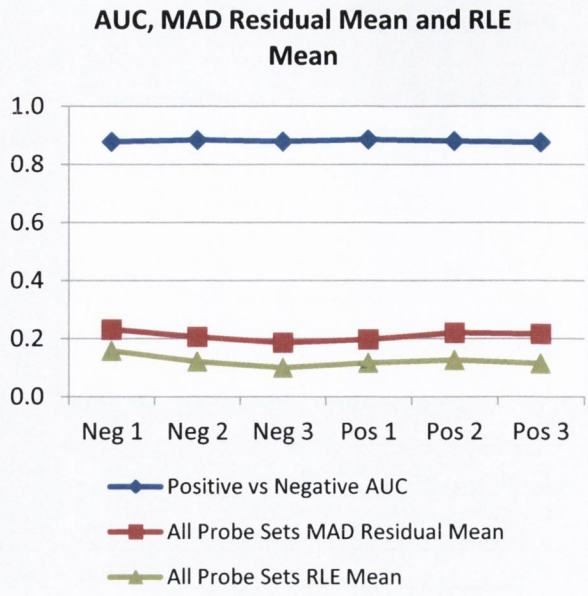
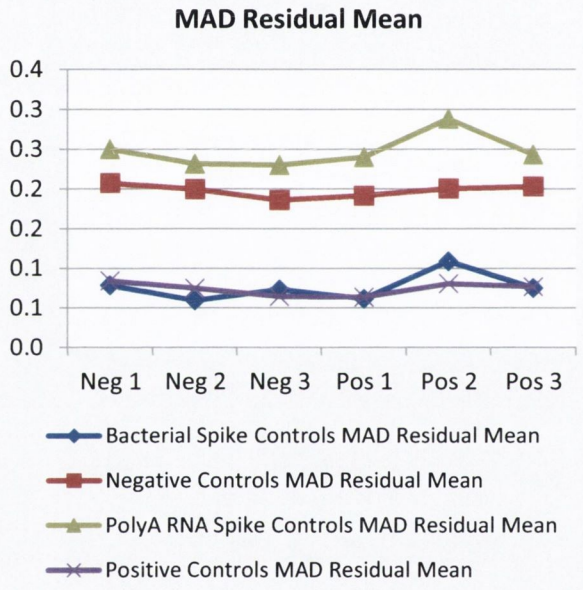
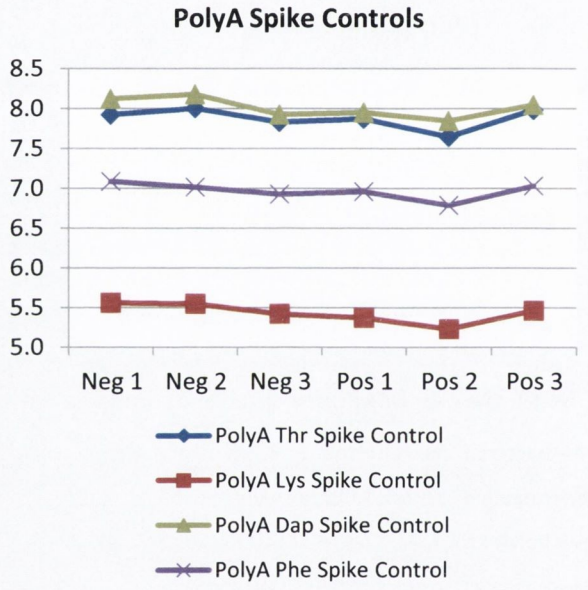


Figure 5.7—Different quality control metrics for gene array analysis of 2102Ep samples. The values on the y-axis are arbitrary. The labels on the x-axis denote each biological replicate of the sample set.









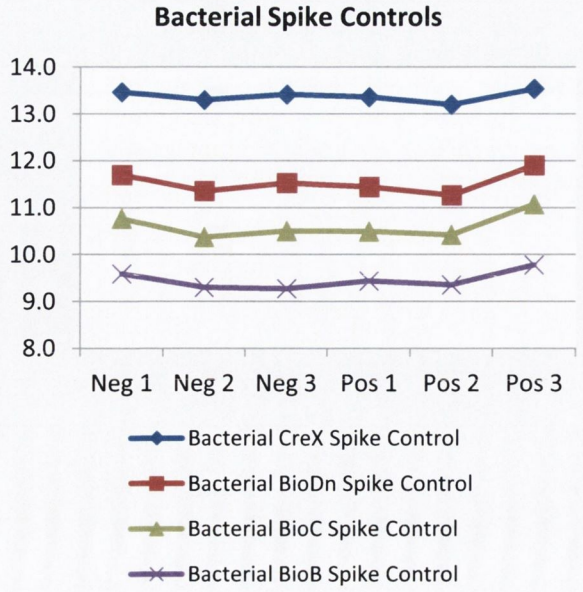


Figure 5.8—Different quality control metrics for gene array analysis of NTera-2 samples. The values on the y-axis are arbitrary. The labels on the x-axis denote each biological replicate of the sample set.

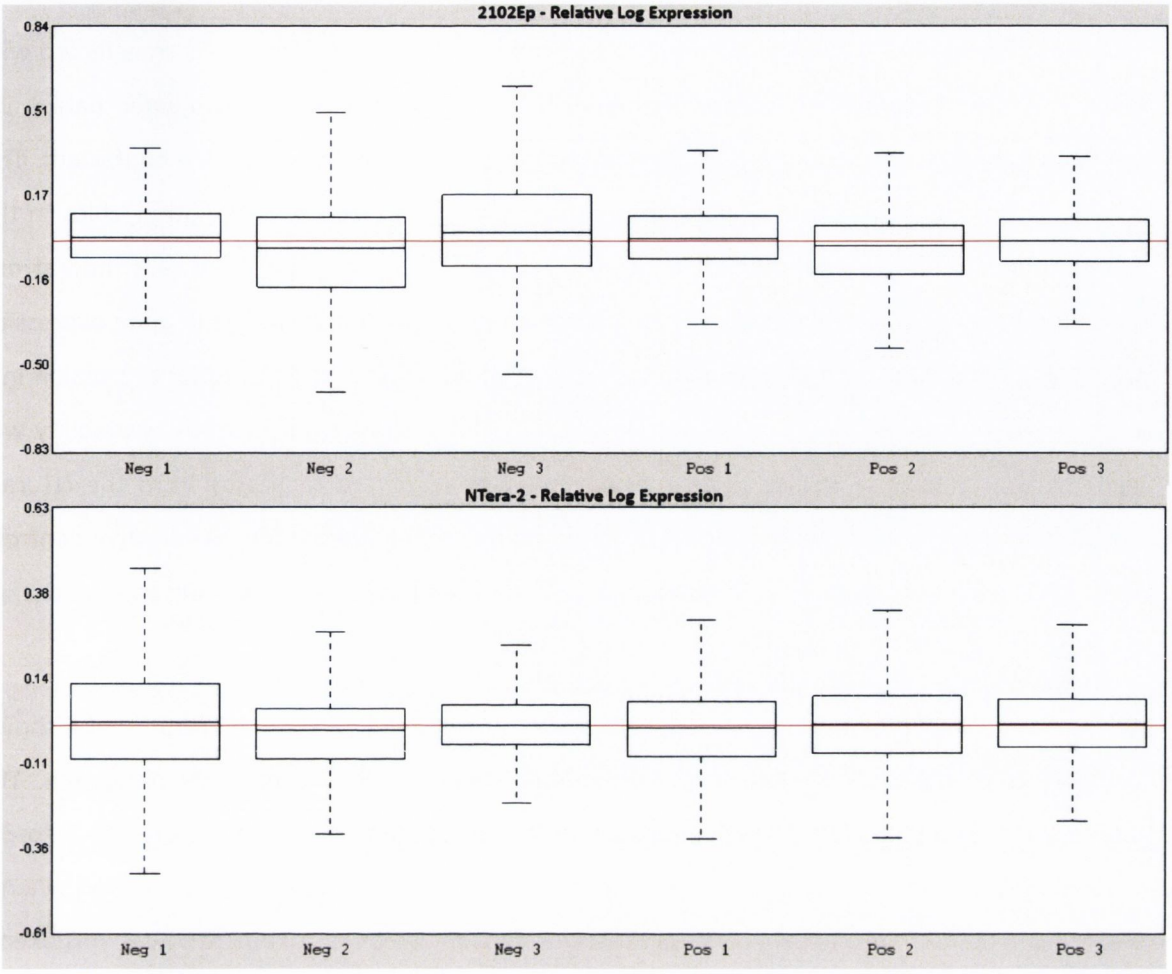


Figure 5.9—Relative Log Expression values in 2102Ep and NTera-2 samples. X-axis values indicate the relative log expression and labels on the y-axis denote each biological replicate of the sample set. The width of the box plots represent the inter-quartile range of the RLE. The middle bar is the RLE median and the error bars represent the limits of the whole range.

5.4.6.2 — Analysis of Differential Gene Expression

Differential gene analysis was performed by comparing 2102Ep or NTera-2 cells transfected with siSOX2 or ssNC#1. Global differential expression was measured with the unadjusted data using the Expression Console software and the Pearson's product-moment correlation coefficient. This correlates the raw expression of each probe on each chip with those of the other chips in the experiment. The results are displayed in Figure 5.11. The results for the 2102Ep cells show strong correlation among all positive controls, preliminarily indicating similar change in gene expression after SOX2 silencing. The high correlation is also indicative of strong differential expression in a large number of genes, while unaltered gene expression is similar across chips. Variability was higher in the 2102Ep negative controls and among some of the replicates from the NTera-2 sample set. Despite this, the positive controls clearly clustered away from the negative controls, although the higher variability among NTera-2 samples might indicate a lower differential change in whole-transcriptome expression compared to 2102Ep samples.

Hiearchical clustering of the normalised, background corrected 2102Ep and NTera-2 data, shown in Figure 5.10 supports the Pearson's correlations described in the previous paragraph. The replicate that had to undergo a repeated sample preparation and hybridisation ('NTera-2 Pos3'), as described in Section 5.4.6.1, appears to be an outlier. This was unfortunately missed when a preliminary differential gene analysis was performed, in which this replicate was mistakenly clustered with its associated replicates. Although this could negatively impact differential gene analysis, highly deregulated genes should still emerge, with few false positives due to the statistical stringency of the LMMA analysis.

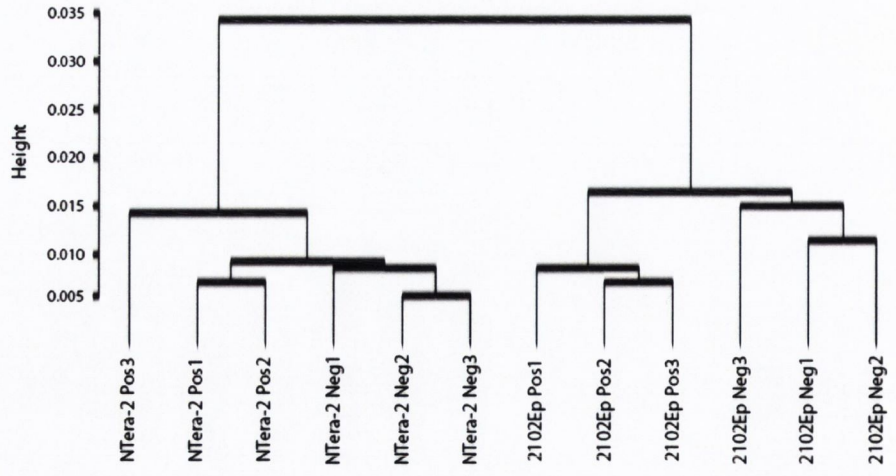


Figure 5.10—Hierarchical clustering of 2102Ep and NTera-2 biological replicates. This dendogram depics the hierarchical clustering of normalised, background-corrected gene expression data of each biological replicate from 2102Ep and NTera-2 cells.

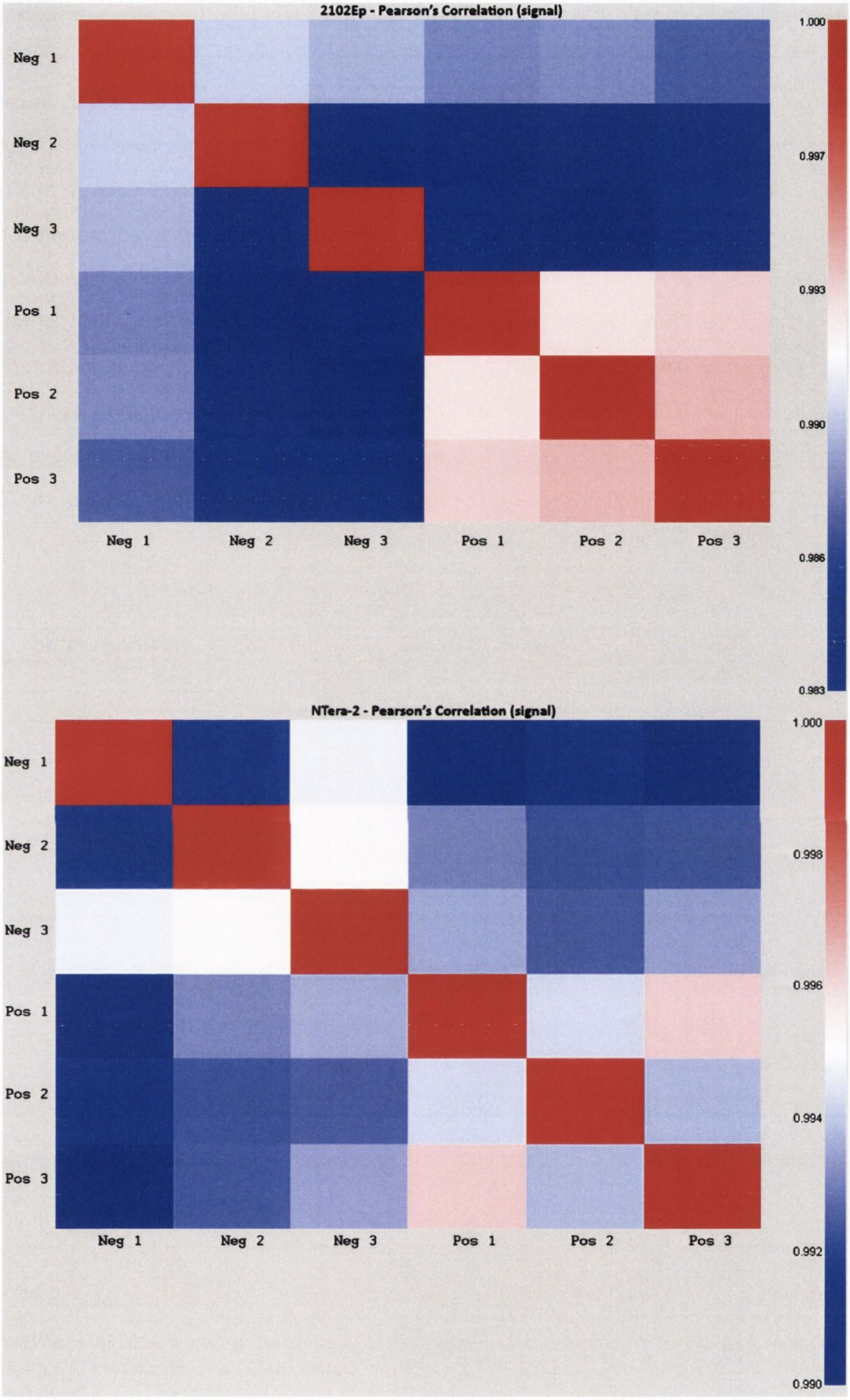


Figure 5.11—Pearson's correlation of unmodified fluorescence. This diagram depicts a summary of the raw, unmodified signal data from each replicate in the experiment. As a heat map it represents all possible pair-wise correlations. (+1 = fully correlated, 0 = fully uncorrolated)

Differential gene expression was computationally calculated with the method described in Section 5.3.2.5. Differentially expressed genes included those that experienced a significant >2-fold expression change, with a $P \leq 0.01$ cut-off limit and an adjusted limit of $P \leq 0.05$ as calculated by the FDR method. In 2102Ep cells, 451 transcripts were differentially expressed as determined by the aforementioned criteria, of which 402 were validated protein coding genes. In NTera-2 cells, 154 transcripts were differentially expressed, of which 131 were validated protein coding genes. The remaining transcripts in both cell lines comprised of a variety of non-coding RNA species such as small nuclear RNA, small nucleolar RNA, Y RNA, long non-coding RNA, rRNA, putatitive genes and other miscellaneous RNAs. This set also contained some intronic and exonic internal controls (see Section 2.5.1.3). These transcripts were not included in further analysis. A summary of the results is shown in Table 5.8. The Venn diagrams in Figure 5.12 highlight the overlap between 2102Ep and NTera-2 differential gene expression. Four genes common between the 2102Ep and NTera-2 samples, were expressed in opposite directions.

Table 5.8—Differentially expressed genes in 2102Ep and NTera-2 cells.

Cell line	Deregulated Genes	Upregulated	Downregulated
2102Ep	402	237	165
NTera-2	131	60	71

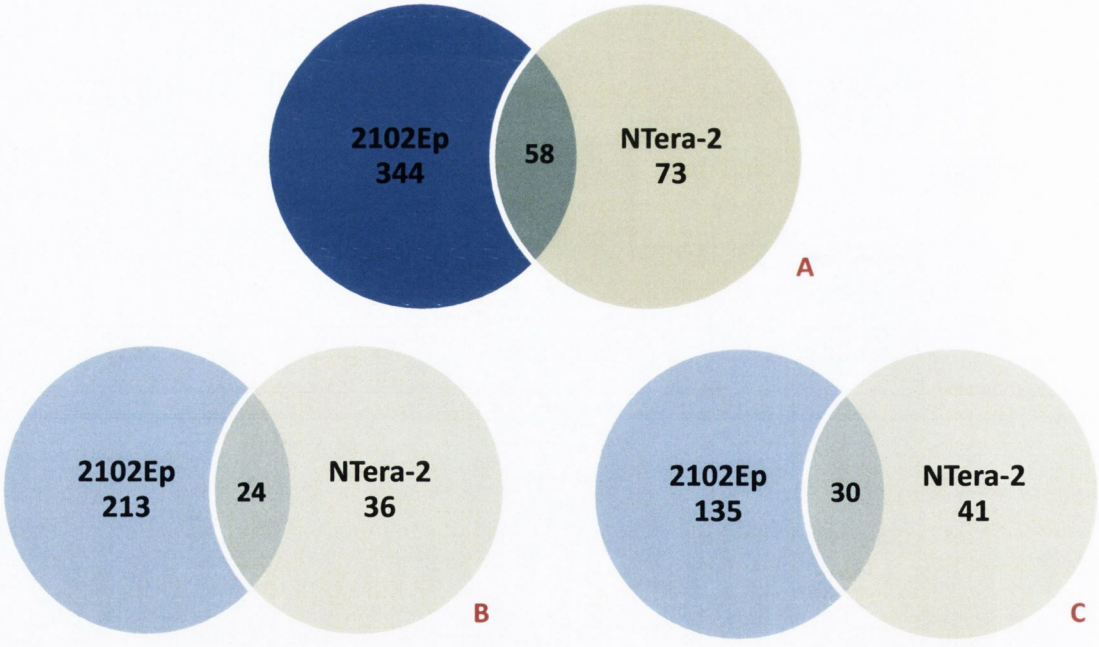


Figure 5.12—Venn diagrams of deregulated genes overlapping between 2102Ep and NTera-2 cells. From all differential genes combined 58 overlap between the 2102Ep and NTera-2 datasets (A). Of these 24 genes are concomitantly upregulated (B) and 30 are concomitantly downregulated (C). Four genes are oppositely deregulated (not shown).

5.4.6.3 — Validation of Gene Array Results by qRT-PCR

QRT-PCR is often considered the 'gold standard' of gene expression analysis, because of its high accuracy and precision. It is also the most common method to validate microarray results.[480] To validate the the gene array expression results, qRT-PCR was used to measure a select number of genes. Gene array data can be validated using two approaches. The first approach is to perform a technical validation of the array data. This will simply test whether the experiment performed on gene arrays is in general agreement with the more representative qRT-PCR results. The second approach is to perform a biological validation, which not only tests the accuracy of the array experiments, but also aims to accurately measure the expression of certain genes of interest to the study. The approach adopted for this study was the former approach, notwithstanding a number of genes measured with qRT-PCR later arose to be very interesting for the discussions in this thesis. With the same RNA samples used for the gene array experiments, the expression of 16 genes was measured in 2102Ep cells, while seven were measured in NTera-2 cells. These were shown to be deregulated >2-fold in each respective cell line. As mentioned in Section 5.3.2.4, the normalisation of the raw gene array data by RMA analysis compresses the differential expression data. This happens especially at the highest and lowest tails of the distribution. To ameliorate the effects from this, the gene array expression results and the results from qRT-PCR were transformed by the natural log. Pearson's product-moment correlation was used to compare gene array and qRT-PCR expression results. Correlation of gene array and qRT-PCR natural log-transformed expression values was high in the 2102Ep data, with significant a Pearson's coefficient of 0.8935. Correlation of the NTera-2 data sets was even higher, with a significant Pearson's coefficient of 0.9697. When the untransformed expression values were compared, some upregulated genes showed a substantial differential increase for the qRT-PCR results, compared to the gene array results. This increase was some cases almost 100 times higher. Although this was similarly the case in NTera-2 cells, the difference between the two expression analyses was much smaller. The gene array and qRT-PCR expression results of the downregulated genes were, relatively, much closer in both the 2102Ep and NTera-2 data sets.

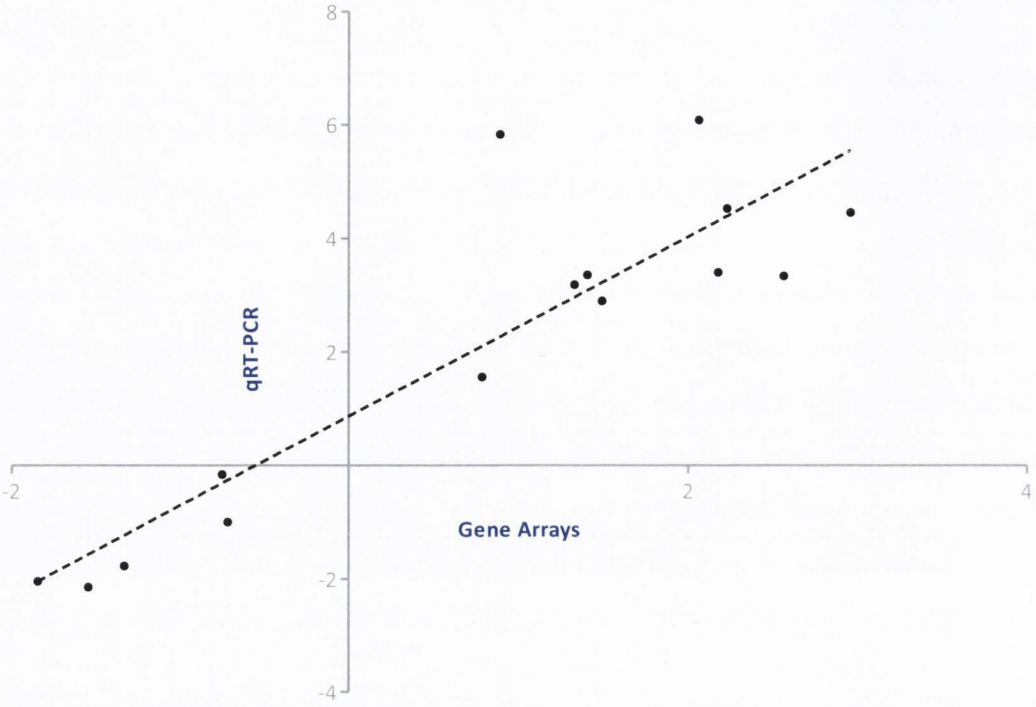


Figure 5.13—Correlation of Ln-transformed gene array and qRT-PCR expression values from 2102Ep cells. Each dot on the graph represents a gene. The dotted line represents the linear regression of whole set. Pearson's correlation coefficient = 0.8935 ($P < 0.0001$). Pearson's product moment correlation coefficient was used to measure significant correlation.

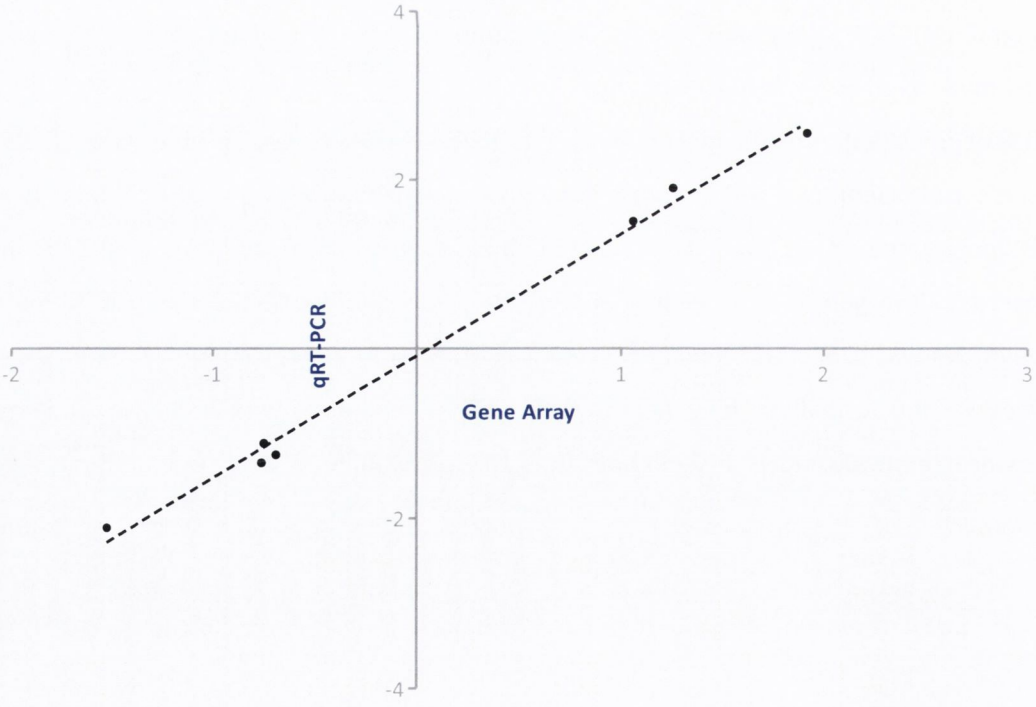


Figure 5.14—Correlation of Ln-transformed gene array and qRT-PCR expression values from NTera-2 cells. Each dot on the graph represents a gene. The dotted line represents the linear regression of whole set. Pearson's correlation coefficient = 0.9697 ($P < 0.0001$). Pearson's product moment correlation coefficient was used to measure significant correlation.

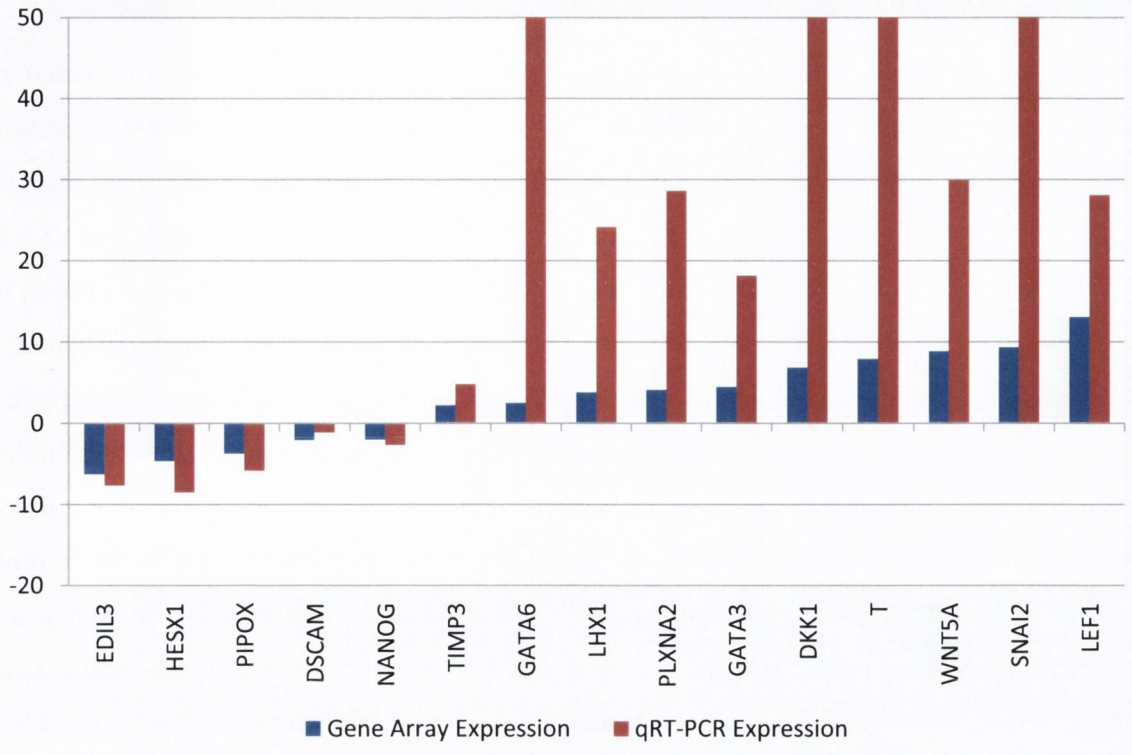


Figure 5.15—The untransformed gene array and qRT-PCR expression values from 2102Ep cells. This chart represents the untransformed fold-expression compared to the non-targeting controls. A cut off was maintained at >50-fold to enable easy visualisation of the expression results. All experiments were performed in biological triplicates.

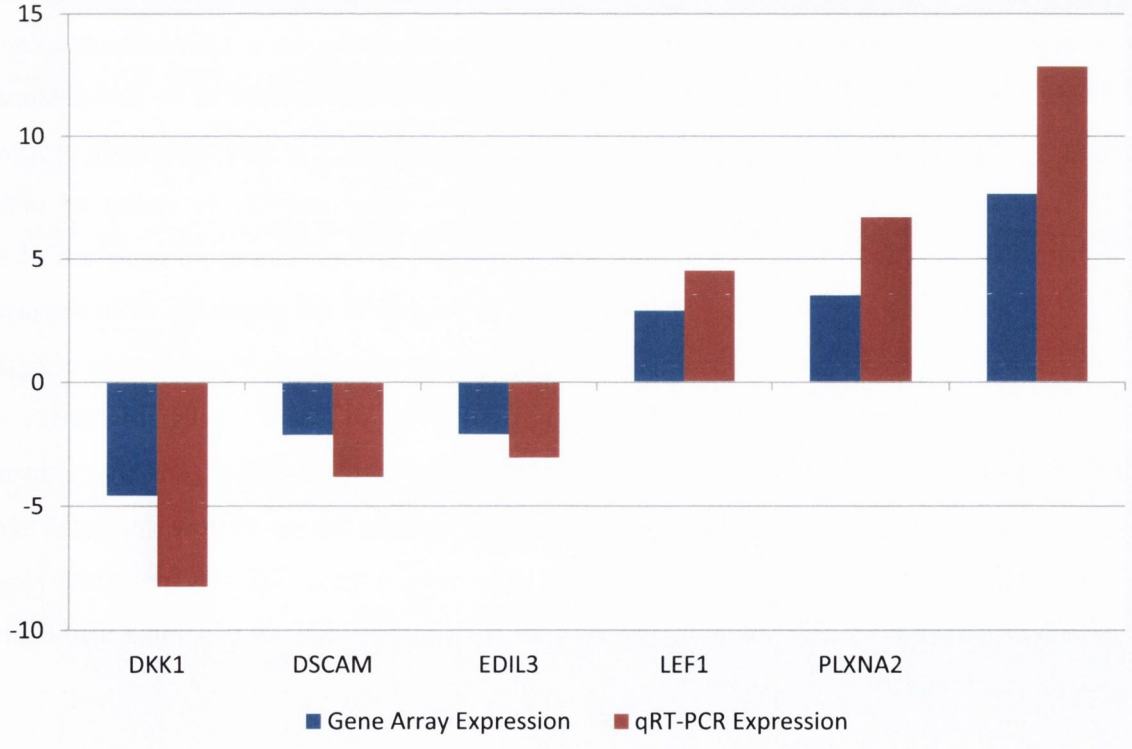


Figure 5.16—The untransformed gene array and qRT-PCR expression values from 2102Ep cells. This chart represents the untransformed fold-expression compared to the non-targeting controls. All experiments were performed in biological triplicates.

5.4.6.4 — Further Inspection of Differential Gene Analysis

The top 10 upregulated and downregulated genes in 2102Ep and NTera-2 cells transfected with siSOX2 are displayed in Table 5.9 and Table 5.10. The full list of genes differentially expressed in both cell lines can be found in Supplemental Table 2. Assuringly, SOX2 was the most downregulated gene in 2102Ep cells. As analysed, in NTera-2 cells, SOX2 was 1.4-fold downregulated and therefore not included in the list of >2-fold deregulated genes. Taking into account the premise that the RMA method of gene array analysis calculates differential expression conservatively, these results were comparable to the gene expression results as measured by qRT-PCR shown in Section 5.4.4. Among the top 10 deregulated genes in 2102Ep cells, notable members are DPPA5 and PRDM14, both of which are important hES and primordial germ cell markers.[335,481–483] Interestingly, in a previous study, DPPA5 was determined to be a central distinguisher between ECCs and hES, embryonic germ cells and primordial germ cells, as it was not in NCCIT and NTera-2 cells.[483] 2102Ep cells were not included in this study, but the expression of DPPA5 and its downregulation by SOX2-induced differentiation shown by the results in Table 5.9 clearly distinguishes these cells from other ECC types. Among the upregulated genes in 2102Ep cells, LGR5 stands out as a crucial marker of both CSCs in intestinal and colon cancer, but also of normal stem cells of the intestine and colon.[484,485] Another highly upregulated gene is MSGN1, a Wnt/β-catenin-activated transcription factor which controls segmentation by activating Notch signalling during early embryogenesis and is upregulated during early ESC differentiation.[486] Similarly, DKK1, HAND1, APLNR, CER1 and MIXL1 have been characterised to be upregulated in differentiating ESCs.[487–490] DKK1, a gene targeted by TCF/β-catenin, has previously been proposed as a high-value target for CSC-directed cancer treatment because of its ability to induce differentiation in CSCs by antagonising the Wnt pathway.[144,145] The strong upregulation of this gene in both 2102Ep and NTera-2 cells (see Table 5.10), supports the possibility of an important SOX2-regulated DKK1-mediated mechanism in the differentiation of ECCs. This is further discussed in Section 5.5.1. Other noteworthy highly upregulated genes in NTera-2 cells are EOMES and HAND1, both markers of early ESC differentiation.[488] Interestingly, GRHL3, a gene that is usually expressed during late embryogenesis, is also upregulated. Among the top 10 downregulated genes in NTera-2 cells, CDH6, a cadherin, has been found to be overexpressed in highly metastatic mouse mammary carcinoma CSCs.[491] Some of the deregulated genes in Table 5.9 and Table 5.10 overlap in both 2102Ep and NTera-2 datasets.

Table 5.11 shows the total 58 differentially expressed genes the two cell lines had in common. Of these, four were oppositely regulated (highlighted in pink). The correlation of the differential expression of all 58 genes between both cell lines is high, with a Pearson's correlation coefficient of 0.8482 ($P < 0.0001$). This correlation is shown in Figure 5.17 and also demonstrates that the genes in 2102Ep cells are almost twice as strongly deregulated than those in NTera-2 cells. When the oppositely expressed genes are removed from the correlation analysis, a Pearson's coefficient of 0.9051 ($P < 0.0001$) is achieved, which indicates a very high correlation of differential expression values. This is a strong indicator that SOX2 regulates these genes in a similar manner in both cell lines.

Finally, as an initial validation of the whole-genome differential expression data, it was assuring to see that several of the genes selected for silencing in Chapter 4, were downregulated in upon SOX2 knock down in 2102Ep and NTera-2 cells. In 2102Ep cells, VSNL1 and AK3L1 (denoted as AK4 in the gene array data) were downregulated. Unexpectedly, DPPA3 and it pseudogene DPPA3P1 were downregulated in NTera-2 cells upon SOX2 knock-down, but this was not found in 2102Ep cells. It is unclear why this is. Besides this gene, AK3L1 and VSNL1 expression was also significantly lower in the NTera-2 gene array data.

Table 5.9—Top 10 upregulated and downregulated genes in 2102Ep cells.

Gene Symbol	Gene Annotation	RefSeq Number	Fold change	FDR-corrected P-value	Boyer et al.	NT2 + ATRA	Fang et al.	Greber et al.
LGR5	leucine-rich repeat containing G protein-coupled receptor 5	NM_003667	28.58	2.17E-10				
MSGN1	mesogenin 1	NM_001105569	19.45	1.84E-09				
DKK1	dickkopf homolog 1 (Xenopus laevis)	NM_012242	19.34	2.39E-09	S O N		↑	
HAND1	heart and neural crest derivatives expressed 1	NM_004821	16.29	2.17E-10	S N			↑
APLNR	apelin receptor	NM_005161	15.17	8.69E-08				
LEF1	lymphoid enhancer-binding factor 1	NM_016269	13.03	2.38E-08	O N	↑		↑
CER1	cerberus 1, cysteine knot superfamily, homolog (Xenopus laevis)	NM_005454	11.37	1.31E-06	S N			
MIXL1	Mix1 homeobox-like 1 (Xenopus laevis)	NM_031944	10.71	3.75E-08				
SNAI2	snail homolog 2 (Drosophila)	NM_003068	9.30	7.72E-07				↑
PRTG	protogenin	NM_173814	9.17	6.85E-08				
SOX2	SRY (sex determining region Y)-box 2	NM_003106	-8.79	2.38E-08	S O N	↑	↓	
IL13RA2	interleukin 13 receptor, alpha 2	NM_000640	-7.49	8.05E-07				
CLEC6A	C-type lectin domain family 6, member A	NM_001007033	-7.48	2.05E-05				
DPPA5	developmental pluripotency associated 5	NM_001025290	-7.08	1.74E-06				
CLEC4D	C-type lectin domain family 4, member D	NM_080387	-6.73	2.23E-06		↓		
OVOS	ovostatin	BX647938	-6.59	9.59E-05				
EDIL3	EGF-like repeats and discoidin I-like domains 3	NM_005711	-6.33	1.60E-05				
NPY1R	neuropeptide Y receptor Y1	NM_000909	-6.12	6.56E-06				
PRDM14	PR domain containing 14	NM_024504	-5.21	1.55E-06	S O N	↓		↓
GOLGA6B	golgin A6 family, member B	NM_018652	-5.06	3.04E-03				

Table 5.10—Top 10 downregulated and upregulated genes in NTera-2 cells.

Gene Symbol	Gene Annotation	RefSeq Number	Fold change	FDR-corrected P-value	Boyer et al.	NT2 + ATRA	Fang et al.	Greber et al.
DKK1	dickkopf homolog 1 (Xenopus laevis)	NM_012242	6.80	4.65E-11	S O N		↑	
SEMA6D	sema domain, transmembrane domain (TM), and cytoplasmic domain, (semaphorin) 6D	NM_153618	4.69	9.32E-09		↑		
DNAJC15	DnaJ (Hsp40) homolog, subfamily C, member 15	NM_013238	4.63	7.67E-07		↑		↑
ALPK2	alpha-kinase 2	NM_052947	4.47	7.07E-08				↑
SNAI2	snail homolog 2 (Drosophila)	NM_003068	4.19	1.15E-07				↑
EOMES	eomesodermin	NM_005442	4.12	5.15E-08	S O N			
GRHL3	grainyhead-like 3 (Drosophila)	NM_198173	3.65	5.20E-09		↑		↑
WLS	wntless homolog (Drosophila)	NM_024911	3.59	7.20E-09		↑		↑
PLXNA2	plexin A2	NM_025179	3.51	8.33E-09		↑		↑
HAND1	heart and neural crest derivatives expressed 1	NM_004821	3.40	1.97E-10	S O N			↑
AK4	adenylate kinase 4	NM_001005353	-7.94	9.61E-09	N	↓		↓
SHISA3	shisa homolog 3 (Xenopus laevis)	NM_001080505	-7.12	1.61E-11		↓		
METTL7A	methyltransferase like 7A	NM_014033	-6.22	7.26E-08	N			↓
TM4SF18	transmembrane 4 L six family member 18	NM_138786	-4.73	1.26E-08		↓		↓
EDIL3	EGF-like repeats and discoidin I-like domains 3	NM_005711	-4.59	4.86E-07				
HHLA1	HERV-H LTR-associating 1	NM_001145095	-4.39	7.42E-08				↓
GOLGA6B	golgin A6 family, member B	NM_018652	-4.38	1.58E-04				
CDH6	cadherin 6, type 2, K-cadherin (fetal kidney)	NM_004932	-3.98	4.93E-10	S	↑		
SEMG1	semenogelin I	NM_003007	-3.71	4.35E-06		↓		
TRIM22	tripartite motif containing 22	NM_006074	-3.40	9.61E-09	S O N			↓

Table 5.11—Genes commonly deregulated in both cell lines.

Gene Symbol	Gene Annotation	RefSeq Number	2102Ep		NTera-2		Boyer et al.	NT2 + ATRA	Fang et al.	Greber et al.
			Fold Change	FDR-corrected P-value	Fold Change	FDR-corrected P-value				
LGR5	leucine-rich repeat containing G protein-coupled receptor 5	NM_003667	28.58	2.17E-10	2.27	8.57E-05				
DKK1	dickkopf homolog 1 (Xenopus laevis)	NM_012242	19.34	2.39E-09	6.80	7.52E-07	SON		↑	
HAND1	heart and neural crest derivatives expressed 1	NM_004821	16.29	2.17E-10	3.40	2.12E-06	SN			↑
APLNR	apelin receptor	NM_005161	15.17	8.69E-08	2.68	1.19E-03				
LEF1	lymphoid enhancer-binding factor 1	NM_016269	13.03	2.38E-08	2.89	1.28E-04	ON	↑		↑
CER1	cerberus 1, cysteine knot superfamily, homolog (Xenopus laevis)	NM_005454	11.37	1.31E-06	2.29	1.44E-02	SN			
MIXL1	Mix1 homeobox-like 1 (Xenopus laevis)	NM_031944	10.71	3.75E-08	2.25	8.11E-04				
SNAI2	snail homolog 2 (Drosophila)	NM_003068	9.30	7.72E-07	4.19	1.20E-04				↑
PRTG	protogenin	NM_173814	9.17	6.85E-08	2.64	2.44E-04				
BMP2	bone morphogenetic protein 2	NM_001200	7.56	6.85E-08	2.24	4.72E-04	SN			
WLS	wntless homolog (Drosophila)	NM_024911	7.34	7.06E-08	3.59	2.39E-05		↑		↑
NTF3	neurotrophin 3	NM_002527	5.73	2.30E-07	2.19	5.88E-04				
EOMES	eomesodermin	NM_005442	4.17	1.31E-05	4.12	7.24E-05	SON			
PLXNA2	plexin A2	NM_025179	4.07	1.42E-06	3.51	2.39E-05		↑		↑
MBNL3	muscleblind-like 3 (Drosophila)	NM_018388	2.95	6.06E-06	2.02	6.76E-04				
FSHR	follicle stimulating hormone receptor	NM_000145	2.92	4.01E-05	2.02	2.99E-03				
CFC1	cripto, FRL-1, cryptic family 1	NM_032545	2.87	7.43E-05	2.28	1.43E-03	N			
ALPK2	alpha-kinase 2	NM_052947	2.82	3.16E-04	4.47	8.57E-05				↑
NTS	neurotensin	NM_006183	2.74	9.92E-04	-2.40	7.33E-03				↓
GAD1	glutamate decarboxylase 1 (brain, 67kDa)	NM_000817	2.72	4.27E-05	2.65	2.25E-04	SN			
NRP1	neuropilin 1	NM_003873	2.52	2.64E-04	-2.22	2.19E-03				
AHNAK	AHNAK nucleoprotein	NM_001620	2.44	1.16E-03	2.58	2.13E-03				
RGS5	regulator of G-protein signaling 5	NM_003617	2.37	5.96E-03	2.53	9.48E-03		↑	↓	↑
ATP2B4	ATPase, Ca++ transporting, plasma membrane 4	NM_001001396	2.32	5.20E-04	2.52	8.11E-04				
CA11	carbonic anhydrase XI	NM_001217	2.28	1.80E-05	2.60	3.10E-05		↓	↓	
DUSP6	dual specificity phosphatase 6	NM_001946	2.12	1.69E-03	-3.17	2.44E-04	SON			
LRRTM1	leucine rich repeat transmembrane neuronal 1	NM_178839	2.07	5.04E-04	2.58	2.44E-04				
TIMP4	TIMP metallopeptidase inhibitor 4	NM_003256	-2.02	1.05E-03	-2.56	4.32E-04	SN			↓
NANOG	Nanog homeobox	NM_024865	-2.04	3.37E-03	-2.17	1.81E-02	SON	↓		
C15orf27	chromosome 15 open reading frame 27	NM_152335	-2.08	6.66E-05	-2.82	2.39E-05				↓
NFE2L3	nuclear factor (erythroid-derived 2)-like 3	NM_004289	-2.08	1.20E-04	-3.07	2.39E-05	SON			↓
SHC3	SHC (Src homology 2 domain containing) transforming protein 3	NM_016848	-2.20	1.64E-03	-2.25	3.82E-03	SON			
CALB1	calbindin 1, 28kDa	NM_004929	-2.21	2.17E-02	2.38	3.15E-02	N			
IQSEC2	IQ motif and Sec7 domain 2	NM_001111125	-2.21	2.90E-04	-2.25	8.08E-04	SN			
TLR4	toll-like receptor 4	NR_024168	-2.29	3.25E-04	-2.35	8.58E-04			↓	
psiTPTE22	TPTE pseudogene	AK097082	-2.32	1.17E-02	-2.55	1.55E-02				↓
PRUNE2	prune homolog 2 (Drosophila)	NM_015225	-2.42	1.83E-03	-2.12	2.44E-04				
C9orf135	chromosome 9 open reading frame 135	NM_001010940	-2.43	1.13E-03	-2.82	1.15E-03			↑	↓
ARSE	arylsulfatase E (chondrodysplasia punctata 1)	NM_000047	-2.52	1.04E-05	-2.50	5.88E-05				↓
VSNL1	visinin-like 1	NM_003385	-2.73	1.37E-02	-2.87	2.61E-02		↓		
SHISA3	shisa homolog 3 (Xenopus laevis)	NM_001080505	-2.76	9.08E-06	-7.12	5.21E-07		↓		

—Table Continued—										
HHLA1	HERV-H LTR-associating 1	NM_001145095	-2.77	3.37E-04	-4.39	8.57E-05				↓
PHC1	polyhomeotic homolog 1 (Drosophila)	NM_004426	-2.79	4.41E-04	-2.24	6.81E-03		↓		↓
METTL7A	methyltransferase like 7A	NM_014033	-3.09	7.64E-04	-6.22	8.57E-05	N			↓
ZNF483	zinc finger protein 483	NM_133464	-3.19	1.73E-05	-2.07	1.83E-03		↓		
AK4	adenylate kinase 4	NM_001005353	-3.25	3.63E-04	-7.94	2.39E-05	N	↓		↓
NR5A2	nuclear receptor subfamily 5, group A, member 2	NM_205860	-3.33	4.83E-05	-2.14	4.48E-03		↓		
PLA2G16	phospholipase A2, group XVI	NM_007069	-3.41	2.54E-06	-2.08	5.88E-04				↓
SEMG1	semenogelin I	NM_003007	-3.47	5.53E-04	-3.71	1.18E-03		↓		
LHFPL4	lipoma HMGIC fusion partner-like 4	NM_198560	-3.63	4.56E-06	-2.37	4.50E-04				
GPR160	G protein-coupled receptor 160	NM_014373	-3.70	1.92E-04	-2.46	7.63E-03				↓
PIPOX	pipecolic acid oxidase	NM_016518	-3.78	1.65E-06	-2.17	4.71E-04	SON			
SAMHD1	SAM domain and HD domain 1	NM_015474	-3.85	1.48E-05	-2.39	1.43E-03		↓		
C10orf96	chromosome 10 open reading frame 96	NM_198515	-4.13	5.30E-04	-3.13	6.86E-03		↓		
GOLGA6B	golgin A6 family, member B	NM_018652	-5.06	3.04E-03	-4.38	1.50E-02				
PRDM14	PR domain containing 14	NM_024504	-5.21	1.55E-06	-2.83	2.55E-04	SON	↓		↓
EDIL3	EGF-like repeats and discoidin I-like domains 3	NM_005711	-6.33	1.60E-05	-4.59	2.66E-04				
OVOS	ovostatin	BX647938	-6.59	9.59E-05	-2.75	1.47E-04				

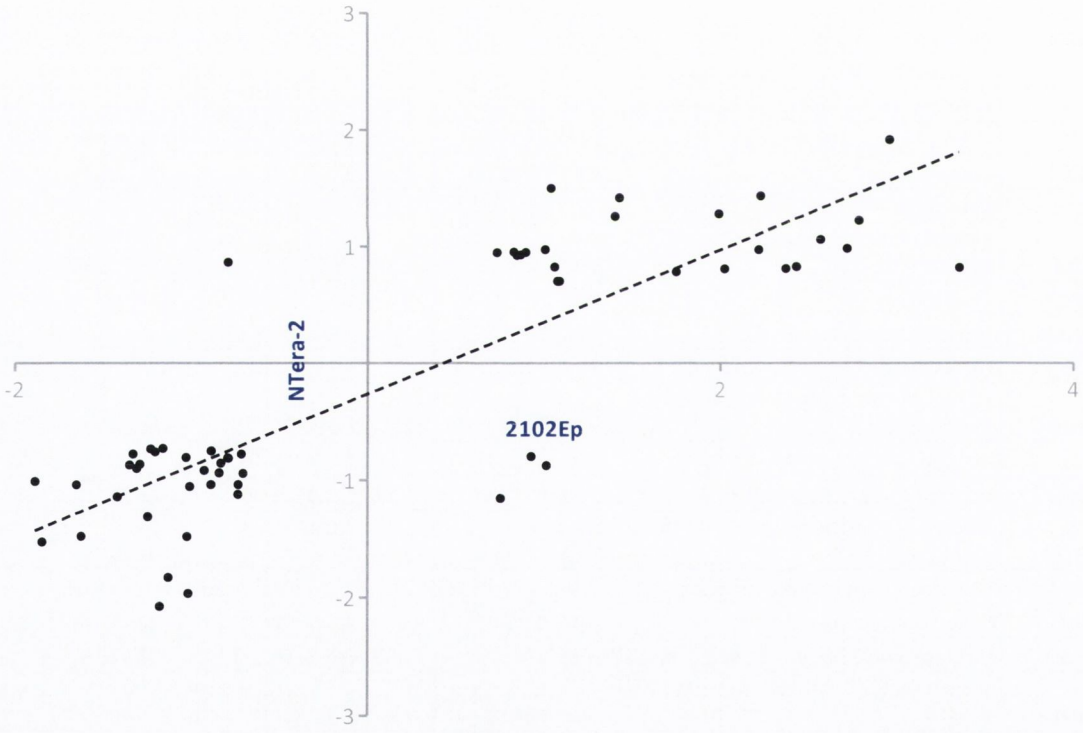


Figure 5.17—Correlation of the 58 differentially expressed genes shared between the 2102Ep and NTera-2 datasets. The axes respresent ln-transformed fold-expression. The dotted line represents the linear regression of the data. Pearson's correlation coefficient = 0.8482 ($P < 0.0001$). Pearson's product moment correlation coefficient was used to measure significant correlation.

5.4.6.5 — Cross-reference with External Studies

The lists of deregulated genes were compared to a number of previously published studies and unpublished data previously produced by this laboratory. To aid in the identification of genes that could be directly regulated in 2102Ep and NTera-2 cells by SOX2 two studies were included in which SOX2 binding sites were identified in hES and GBM cells using ChIP sequencing (ChIP-seq).[79,466] Although these cell lines are different than the ones used in this thesis, the previous establishement of *in vivo* promoter binding sites of SOX2 in other cell types, increases the probability that genes found in the data from this thesis that have previously found SOX2 binding sites, are regulated by this transcription factor in the cell lines used in this study. Additionally, the data was compared to a previous study in which early differentiation of NCCIT cells was induced by SOX2 silencing.[378] Finally a comparison was also made with previous data from this laboratory containing whole-genome expression results during early differentiation of NTera-2 cells by ATRA. The purpose of the last two studies was to aid with the identification of potentially important genes for ECC differentiation. A comparison of the data from these studies and the differential gene expression data produced for this chapter is added to Table 5.9, Table 5.10 and Table 5.11. The studies used for comparison are the following:

1. With ChIP, Boyer et al. (2005) performed a genome-wide search for promoter binding sites of the core pluripotency transcription factors, SOX2, OCT4 and NANOG, in hESCs. This showed that these factors were bound to the promoter regions belonging to thousands of genes.[79]
2. As mentioned in Section 5.4.6.2, a genome-wide differential expression analysis was previously performed by this laboratory with NTera-2 cells in 10^{-2} mM ATRA for three days. This produced hundreds of differentially regulated genes that had functions in the retinoid pathway, but also in early NTera-2 differentiation.
3. In a similar study to this one, Fang et al. (2011) conducted a functional study by silencing SOX2 in a GBM cell line and conducting a genome-wide expression and ChIP-seq analysis, three days after SOX2 silencing.[466] Differential gene expression results and ChIP-seq results were included in the comparison with the data in this Section.
4. As mentioned in Section 4.5, SOX2 was previously silenced in NCCIT cells, a different ECC line from 2102Ep and NTera-2 cells by Greber et al (2007).[378] Two days after SOX2 silencing, whole-genome differential expression analysis was performed.

For study 1, the promoter-bound pluripotency factors are listed in Table 5.9, Table 5.10 and Table 5.11 as S, O, N for SOX2, OCT4 and NANOG resepectively. Study 2 is listed as 'NTera-2 + ATRA' in these tables, all other studies by their respective first authors. For studies 2-4, the genes which

overlap with those in Table 5.9, Table 5.10 and Table 5.11 are denominated with an arrow in concordance to their regulation in the compared study. Genes that were found to be bound by SOX2 in study 3 were marked in the tables with grey.

A large degree of overlap exists between the differentially expressed genes in this study and the genes bound by OCT4, SOX2 and NANOG the study by Boyer et al. (2005). Table 5.12 shows a further breakdown of the numbers of differentially regulated genes in this chapter which are found Boyer et al. (2005) data and, thus, have a demonstrated capability of being bound by SOX2, OCT4 and NANOG. The percentage value in brackets represents the overlapping data as a fraction of each group of total, upregulated or downregulated genes. Of all differentially expressed genes in 2102Ep cells, 12.4% are found to be bound by SOX2 in hESCs. In NTera-2 cells, 21.4% of differentially expressed genes are bound by SOX2 in hESCs.

Table 5.12—Numbers of genes bound by OCT4/SOX2/NANOG as per Boyer et al. with the percentage fraction of each respective group.

	Bound by OCT4/SOX2/NANOG			Upregulated			Downregulated		
	Total	Up	Down	SOX2	OCT4	NANOG	SOX2	OCT4	NANOG
2102Ep	95 (24%)	55 (23%)	40 (23%)	25	22	51	25	17	38
NTera-2	42 (32%)	17 (28%)	25 (35%)	9	7	17	19	14	23

In Table 5.13, the number of genes common with the data from the studies by Greber et al. (2007) and the study conducted previously by this laboratory are listed. A higher overlap exists between the NTera-2 differential gene expression from this Section and the data from the 'NTera-2 + ATRA' study, than when the 2102Ep data is compared with the latter study. This is expected, as the differentiation program as induced by ATRA in NTera-2 cells is probably more comparable to the differentiation program as induced by SOX2 in NTera-2 cells, than in 2102Ep cells. The comparison with the NCCIT SOX2 knock-down data from Greber et al. (2007) shows reasonable overlap of differential gene expression in this embryonal carcinoma cell line when compared with 2102Ep and NTera-2 cells. However, overall, this overlap is higher in NTera-2 cells than in 2102Ep cells, possibly indicating that the differentiation program or SOX2-mediated gene regulation in NCCIT cells is more comparable to NTera-2 cells than 2102Ep cells. This has not yet been suggested and might indicate a significant difference in the biology of different ECC types.

Table 5.13—Numbers of genes that overlap with the data from Greber et al and this laboratory with the percentage fraction of each respective group.

	NTera-2 + ATRA			Greber et al.		
	Up	Down	Opp.	Up	Down	Opp.
2102Ep	27 (11%)	30 (18%)	10 (2%)	57 (24%)	33 (20%)	2 (0%)
NTera-2	10 (17%)	24 (34%)	3 (2%)	23 (38%)	23 (32%)	1 (1%)

The number of genes that overlap with the differential expression data from the GBM SOX2 knock-down study by Fang et al. (2011) are listed in Table 5.14 and show surprisingly little overlap. Contrarily, a much larger proportion of differentially expressed genes in 2102Ep and NTera-2 cells are bound by SOX2 in GBM cells according to the Fang et al. (2011) study. This probably indicates that SOX2 could have quite dissimilar functions in GBM, than in embryonal carcinoma. The comparisons of the data with all aforementioned studies are further discussed in the discussion of this chapter.

Table 5.14—Numbers of genes that overlap with the data from Fang et al. with the percentage fraction of each respective group.

	Gene Expression			SOX2 ChIP-seq		
	Up	Down	Opp.	Total	Up	Down
2102Ep	6 (3%)	6 (4%)	8 (2%)	93 (23%)	55 (23%)	38 (23%)
NTera-2	2 (3%)	1 (1%)	7 (5%)	39 (30%)	19 (32%)	20 (28%)

5.4.6.6 — Pathway Analysis

Various deregulated genes in 2102Ep and NTera-2 cells are members of particular families of cellular pathways. Categorising the deregulated genes into these pathways may clarify the biology of the SOX2-induced differentiation of these cell lines. With the aid of the online DAVID Bioinformatic Resources 6.7 software (described in Section 5.3.2.6) in combination with the online Kyoto Encyclopedia of Genes and Genomes (KEGG) PATHWAY mapping module, the differentially regulated genes were sorted into various cellular pathways.[492] The DAVID software cross-referenced the input genes from each cell line with the pathway database from KEGG PATHWAY. The genes were categorised according to which pathway they belonged. Because cellular pathways frequently cross-link, various genes were categorised in multiple pathways. The pathways found to be highly represented in the data have historically been highly associated with stemness, embryonic development and cancer. These are Wnt, MAPK, TGFβ, Hedgehog, JAK/STAT, PPAR and chemokine signalling. Figure 5.18 and Figure 5.19 depict the member genes of different pathways and their fold expression in NTera-2 and 2102Ep cells respectively. The results show that, particularly in 2102Ep cells, the expression of many signalling pathway components are activated by SOX2 downregulation. The importance of these genes and there potential contribution to 2102Ep and NTera-2 cell carcinogenicity and stemness are further detailed in the discussion of this chapter.

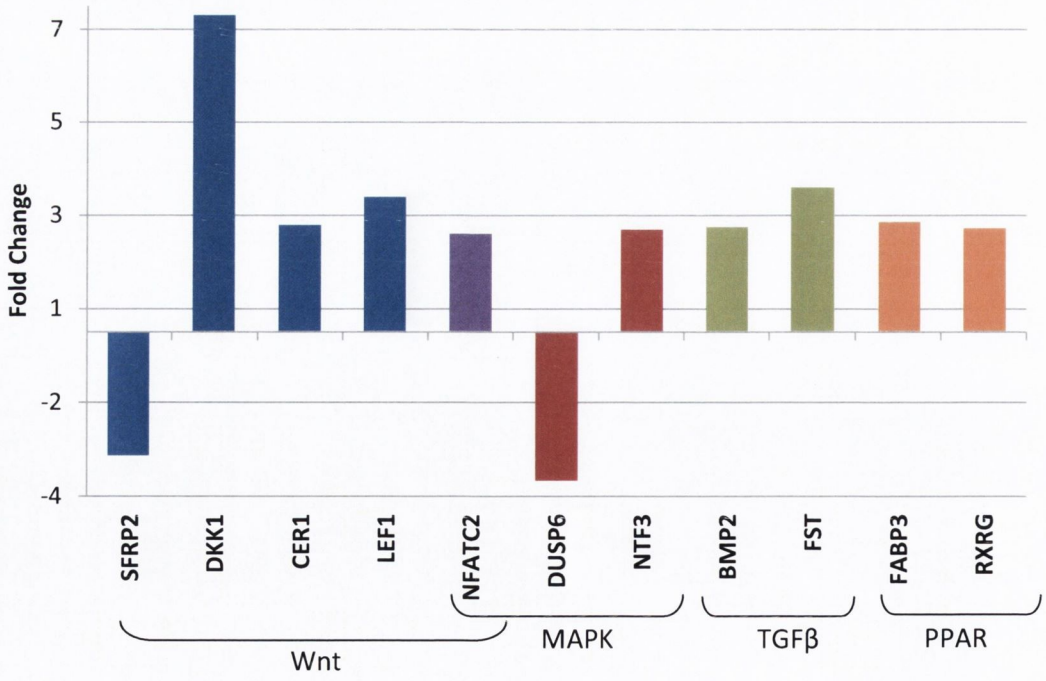


Figure 5.18—Deregulated genes in important signalling pathways in NTera-2 cells. The horizontal brackets encompass the listed genes in the associated pathway. The genes at points of overlap are simultaneously categorised in two different pathways.

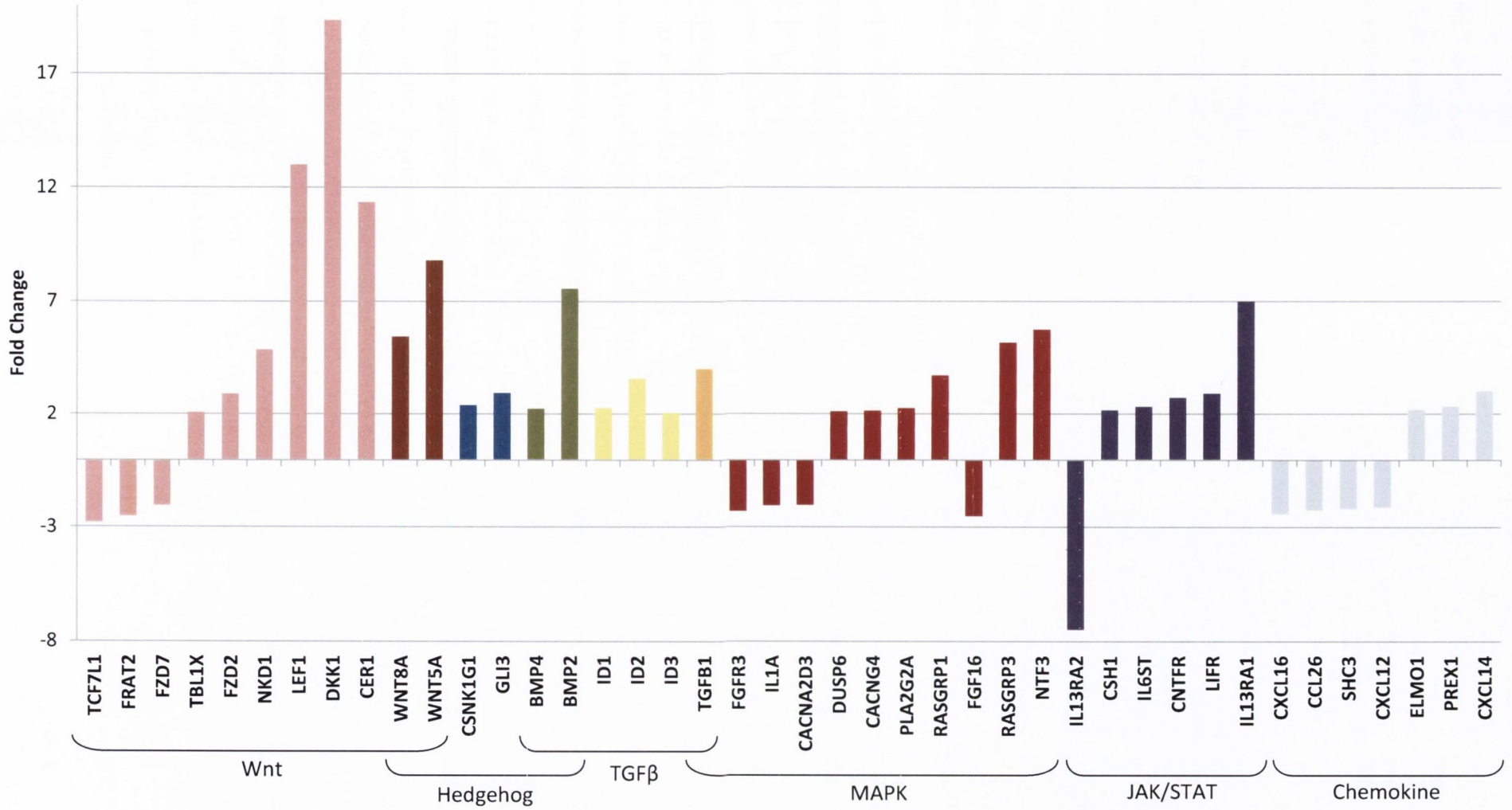


Figure 5.19—Deregulated genes in important signalling pathways in 2102Ep cells. The horizontal brackets encompass the listed genes in the associated pathway. The genes at points of overlap are simultaneously categorised in two different pathways.

5.4.6.7 — Functional Gene and Annotation Analysis

Functional annotation analysis was performed with DAVID Bioinformatic Resources 6.7 for all differentially expressed genes in 2102Ep and NTera-2 cells as described in Section 5.3.2.6. Functional annotation analysis does not provide a very detailed and reductionistic description of the molecular biology in the relevant cell processes, but instead provides a more holistic method of broadly analysing and comparing cell phenotype. Upregulated and downregulated genes were separately analysed. In this Section, the functional annotation clusters were depicted as circle graphs with a description that summarises the most dominant annotations in the cluster (see Figure 5.20 and Figure 5.21). Because classification stringency was set quite high, the clusters describe quite specific processes. The circle graphs represent only the significantly enriched functional annotation clusters, with a maximum of top 10 ranked by Group Enrichment Score, an indicator of significance (higher is more significantly enriched). A detailed description of the functional annotation clusters with a break-down of individual terms are found in Supplemental Table 3.

The functional annotation results from the upregulated genes in 2102Ep cells, shown in Figure 5.20, are highly enriched for terms that relate to embryonic morphogenesis, such as Mesenchyme Development, Embryonic Limb Development, Pattern Formation and Regionalisation and Axis, Embryonic Pattern Specification and Homeobox. This strongly indicates that the induced downregulation of SOX2 activates processes involved in ESC differentiation and embryonic morphogenesis. Of particular interest is the upregulation of HOX genes, a common feature of early embryonic development. The Homeobox term includes about 12 genes, which are part of a large superfamily, containing transcription factors with a homeobox motif. The HOX network is frequently associated with developmental fate, but, more importantly for this study, have been frequently associated with the differentiation of various ECC lines.[493–496] The large majority of genes in the annotation cluster, Regulation of Transcription, are transcription factors and represent the largest annotated group in the dataset. Almost all members of this annotation cluster overlap with the second largest cluster, Positive Regulation of Transcription and Macromolecule Biosynthesis. Both groups contain a large number of ESC markers. With a total of 55 unique genes in both annotation clusters, this indicates a substantial activation of transcriptional networks, a feature of ESC differentiation. A number upregulated genes are members of the Mesenchyme Development and Cell Migration and Motility. These are functions that are frequently associated with ES differentiation and early embryonic development. However, they have also commonly been implicated with oncogenicity and metastasis.

Among the downregulated genes far less prominent annotation enrichment is found than among the upregulated genes. The most significantly enriched cluster, as defined by its Group Enrichment score contain is High Mobility Group (HMG), a family of genes of which SOX2 is a member. This family is involved in transcriptional regulation by the structural modification of chromatin superstructures.[497] This signifies the regulation of other HMG genes by SOX2, such as SOX15, TOX3, WDHD1 and TCF7L1. Together with SOX2, SOX15, TOX3 and TCF7L1 are important pluripotency markers, but also have been classified oncogenes in germ cell tumours and CSCs.[86,498–500] Strikingly, 13 genes belonging to the cluster, Regulation of Cell Death, are downregulated in 2102Ep cells. This is the largest significantly enriched group and contains many pro-apoptosis genes (BNIP3, EPHA7, EDNRB, IL1A, PRUNE2, TERF1 and CD24) which appears to suggest that SOX2-silenced 2102Ep cells may be becoming more resistant to apoptosis.[501–507] CD24 is of particular interest as the low expression of this cell surface protein, in combination with high CD44 expression, is a common CSC marker.[2,485,508] The upregulation of CD24 in breast CSCs has furthermore resulted in the apoptosis of these cells.[507] However a number of very important oncogenic anti-apoptotic genes are also downregulated (TDGF1, PIM2 and TLR4).[509–511] Another gene of interest is TDGF1, as the upregulation of these gene has identified a highly tumorigenic subpopulation of in the NTera-2 ECC line. It is therefore possible that this gene might have a similar oncogenic role in 2102Ep cells. Also highly interesting is the downregulation of TLR4 as this gene is an important marker of ovarian CSCs with functional implications in the maintenance of these cells.[512]

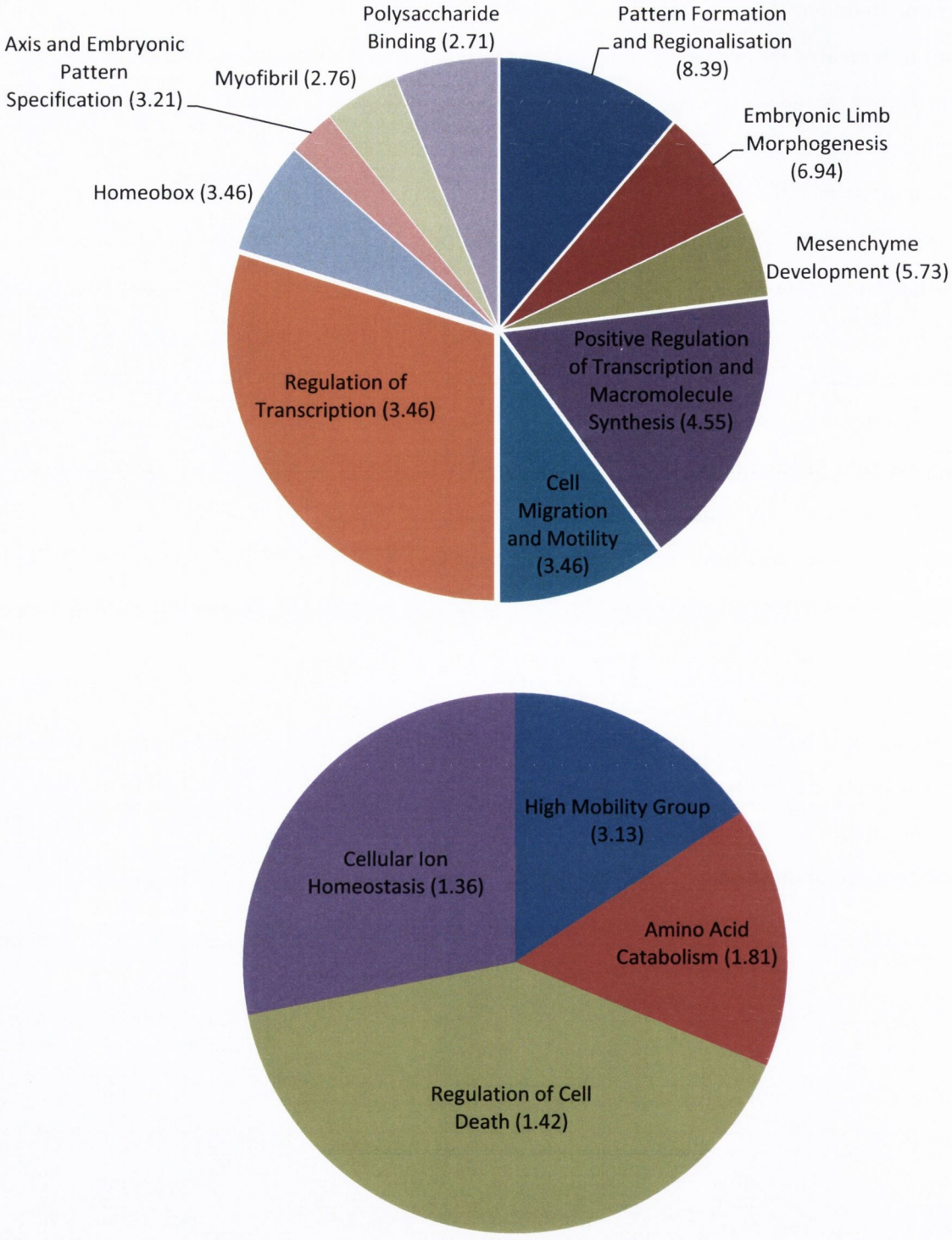


Figure 5.20—Most highly enriched functional annotation clusters in 2102Ep differential gene expression data. The top circle graph depicts functional annotation enrichment in upregulated genes and the bottom circle graph represents the downregulated genes. Each segment is labelled with the dominant term of the related functional annotation cluster and its size is proportional to the number of genes in the cluster as a fraction of the whole group. The number in brackets is the Group Enrichment Score of the cluster; higher indicates more significantly enriched.

The list of deregulated genes in NTera-2 cells shows less overall annotation enrichment than the data from 2102Ep cells. Among the upregulated genes, several signficantly enriched functional annotation clusters, shown in Figure 5.21, describe embryonic development and germ layer formation (Trophectodermal/Blastocyst Formation and Embryonic Development, Mesenchymal Develoment and EMT, Primary Germ Layer and Mesoderm Formation, Uretic Bud/Tube/Epithelium Development, Determination of Symmetry). However, the total number of unique genes in these clusters (16) is relatively low in comparison to the size of the entire dataset. By far the largest and most enriched annotation cluster is described by the very general term 'Signal' and was excluded from Figure 5.21 for the purpose of visualising more specialised clusters. An interesting cluster is 'Mesenchymal Develoment and EMT' as this contains three genes crucial to EMT (BMP2, EOMES and LEF1). All three these genes aid in the induction of EMT, a process which is elemental embryonic development, but is also strongly associated with cancer cell migration and CSC formation.[513–515] Oddly not included in the aforementioned cluster by DAVID, but upregulated in NTera-2 cells is also SNAI2/Slug, a crucial inducer of EMT and repressor of E-cadherin.[111,516] Unexpectedly CDH1/E-cadherin is upregulated in NTera-2 cells despite the upregulation of pro-EMT genes. This finding is further discussed in Section 5.5.1.2.

Finally, among the downregulated genes in NTera-2 cells, only one functional annotation cluster is significantly enriched (not shown as a circle graph). This describes Cell/Neuron Projection Morphogenesis and Neuronal Differentiation. Of the six genes in this cluster, DSCAM, KITLG, ADM, FEZ1 and NRP1 have been directly associated with functional roles in neural development and, in some cases, neural stem cells.[517–520] KITLG is also involved in germ cell development and has also been described as an oncogene, particularly in testicular germ cell cancer, while FEZ1 has been strongly associated with ovarian cancer.[521,522]

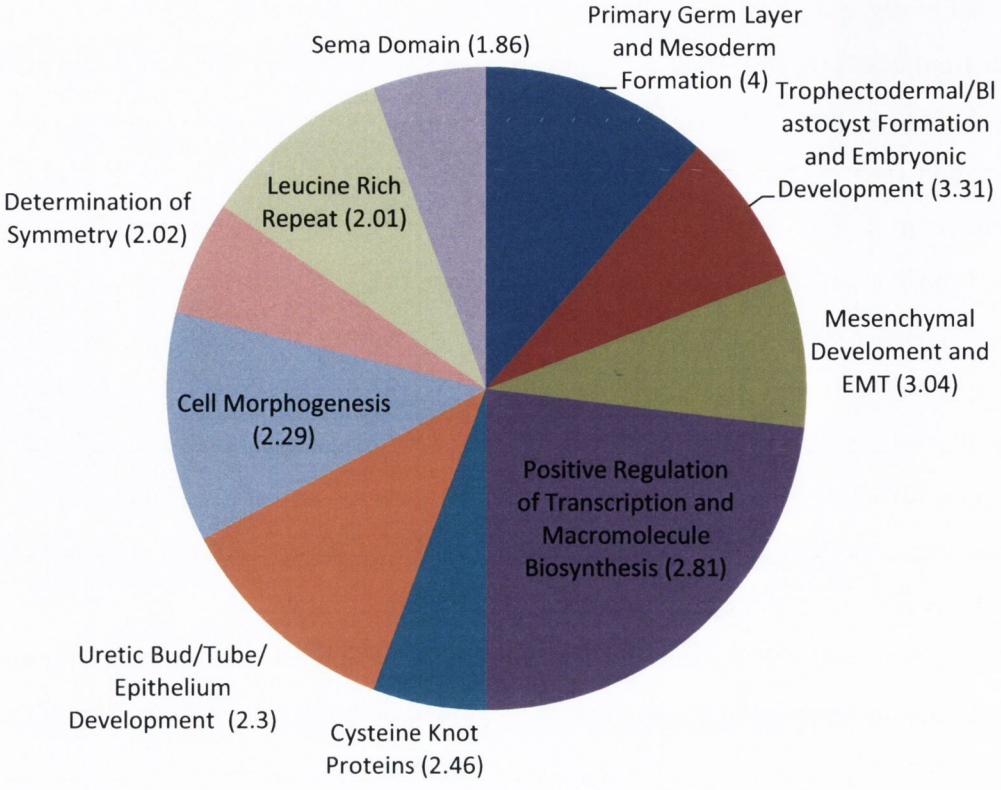


Figure 5.21—Most highly enriched functional annotation clusters among upregulated genes in NTera-2 cells. The circle graph depicts functional annotation enrichment in upregulated genes Each segment is labelled with the dominant term of the related functional annotation cluster and its size is proportional to the number of genes in the cluster as a fraction of the whole group. The number in brackets is the Group Enrichment Score of the cluster; higher indicates more significantly enriched.

5.4.7- Whole miRNA Expression Analysis with TaqMan® Arrays

For the analysis of whole-microRNAome expression in 2102Ep and NTera-2 cells after siSOX2 silencing, TaqMan® Array Human MicroRNA A+B ver. 3.0 cards were used. As described in Section 2.5.1.2, these cards provided the means to measure the differential expression of up to 754 miRNAs in both cell lines.

5.4.7.1 — Optimisation of the miRNA Preamplification Stage

Applied Biosystems recommends a preamplification of miRNA in the original RNA sample to enhance the performance of the TaqMan® Arrays. This is especially recommended when RNA concentrations are low (i.e. approximately 30 ng/µl). The concentrations of the RNA samples used in this experiment are much higher (i.e. approximately 300 ng/µl). An experiment was conducted to decide whether preamplification was desirable and consistent. MiRNA cDNA was synthesised from 210 ng total RNA and a part preamplified and diluted 1:4 in 0.1X TE buffer according to the protocol in Section 5.3.3.1. Unamplified and preamplified cDNA was assayed for miR-16 expression. The results for NTera-2 cells can be found in Figure 5.22. Unamplified samples show a miR-16 expression at 30 Ct, while miR-16 expression in the preamplified sample is at 15.8 Ct. In conventional qRT-PCR experiments conducted during this study, miR-16 expressed within the 18-22 Ct range. Considering the substantially higher concentration of RNA used in this experiment, the expression of miR-16 in the preamplified cDNA reflects the expected range closer than the unamplified cDNA. When taking into account the maximum possible amount of RNA which can used for cDNA synthesis (approximately 800 ng), miR-16 would express at a predicted 28 Ct. This was considered inadequate for the TaqMan Array®, so preamplification was the chosen option for further experimentation.

The expression results from 2102Ep samples showed high variance as seen in Figure 5.23. After a repeat of this experiment (data not shown), the 2102Ep RNA was concluded to probably contain inhibitory factors which interfered with the cDNA synthesis. As this was not detected before in previous qRT-PCR experiments, this anomaly is probably limited to qRT-PCR with the MegaPlex™ Primer Pool. To find an ideal RNA concentration at which inhibitory effects could be minimised a range of RNA dilutions underwent cDNA synthesis, preamplification and qRT-PCR. The range of dilutions and the results of miR-16 expression are displayed in Figure 5.24. MiR-16 expression peaked at 105 ng RNA. This was chosen as the quantity for further experimentation.

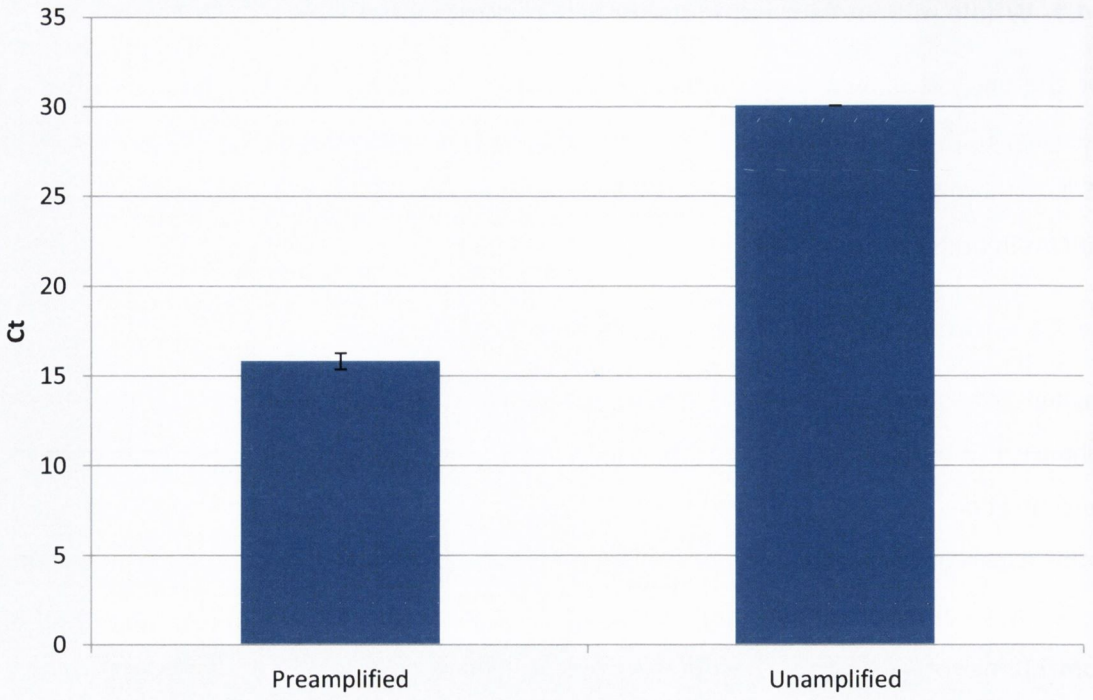


Figure 5.22—MiR-16 expression in NTera-2 cells using unamplified an preamplified cDNA. The Ct value in preamplified cDNA is significantly lower (i.e. expression is significantly higher) in preamplified cDNA compared to unamplified cDNA ($P < 0.0001$). Values are normalised mean expression ± SD of biological triplicates. Sample means were compared with a t-test.

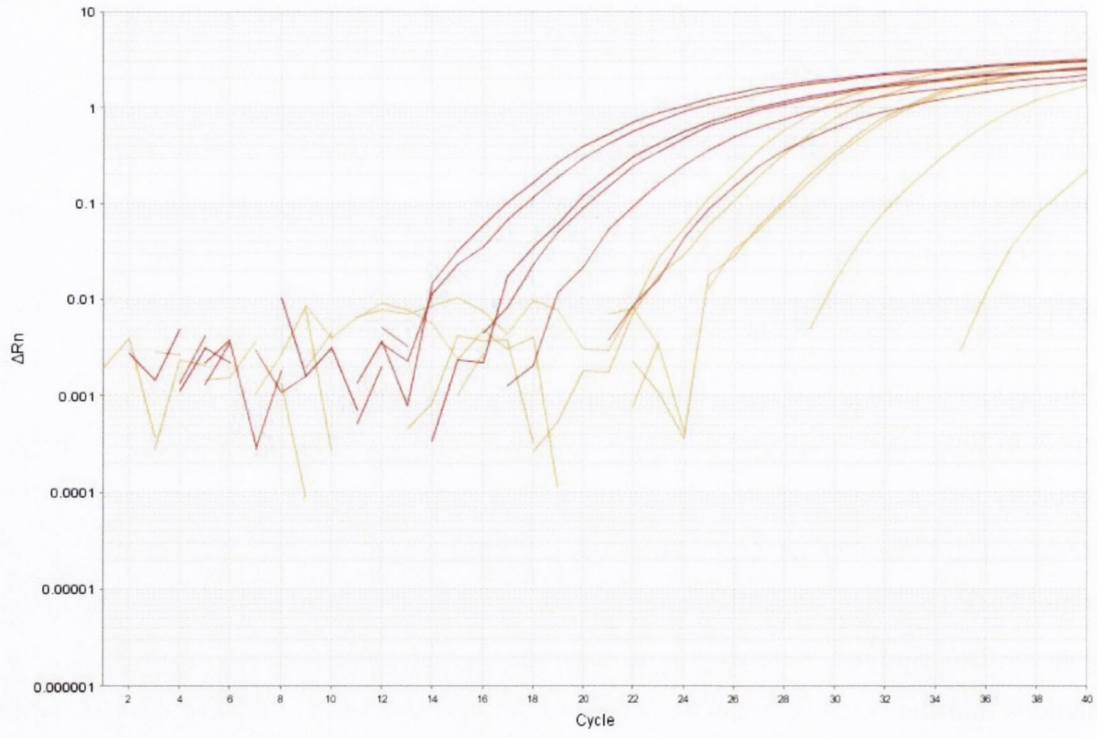


Figure 5.23—qRT-PCR amplification plot of miR-16 expression in preamplified and unamplified 2102Ep cDNA. Red lines represent miR-16 expression in preamplified cDNA, yellow lines in unamplified cDNA. The values on the x-axis represent the qRT-PCR cycles.

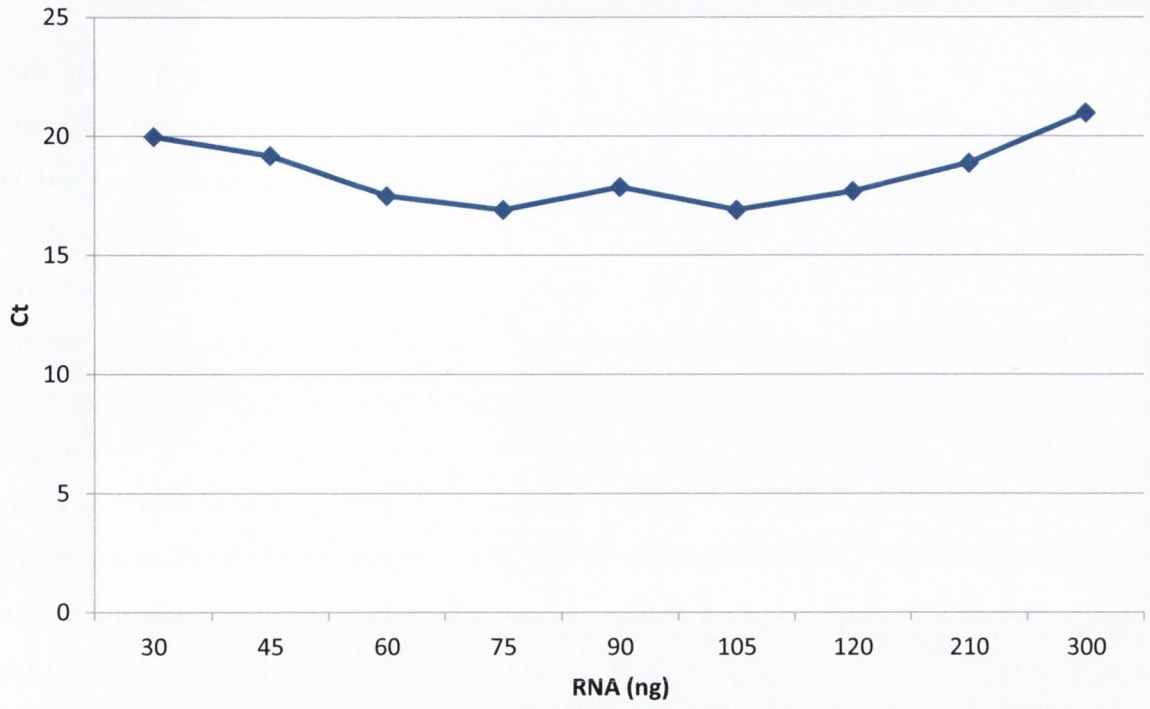


Figure 5.24—MiR-16 expression in cDNA synthesised and preamplified from different quantities of 2102Ep RNA. The lowest Ct value (i.e. highest expression) for miR-16 is at 105 ng RNA.

5.4.7.2 — Quality Control

The protocol for quality control is detailed in Section 5.3.2.3. Following qRT-PCR and the thresholds and baselines of each amplification plot was either automatically or manually set with the RQ Manager v1.2.1 software *(Applied Biosystems)*. MiRNAs with no expression or with highly aberrant amplification plots were omitted from the overall data set. The processed data was imported in the DataAssist v3.0 relative quantification (RQ) software *(Applied Biosystems)*. This tool was used for producing differential expression results and for quality control. Two quality control metrics were used, overall range of Ct distribution and signal correlation between individual samples. In Figure 5.26 box plots of the of Ct distribution gamut is shown for each sample. The box plots show that mean and median Ct of all samples keep within <1 Ct, which indicates low variance between the samples. This also indicates the difference in specific miRNA expression between both control groups is probably subtle in terms of quantity and amplitude. 50% of all expression lies within a 21-28 range, which is the desired range for high quality expression results. Signal correlation as revealed in Figure 5.25 and Figure 5.27 for 2102Ep and NTera-2 cells respectively, show high correlation among the biological replicates in the control groups, although NTera-2 B card data shows markedly less replicate correlation compared to the other data sets. In addition, 2102Ep Neg1 of the A card data shows less correlation than the other replicates in the sample set, despite good correlation of the other ones.

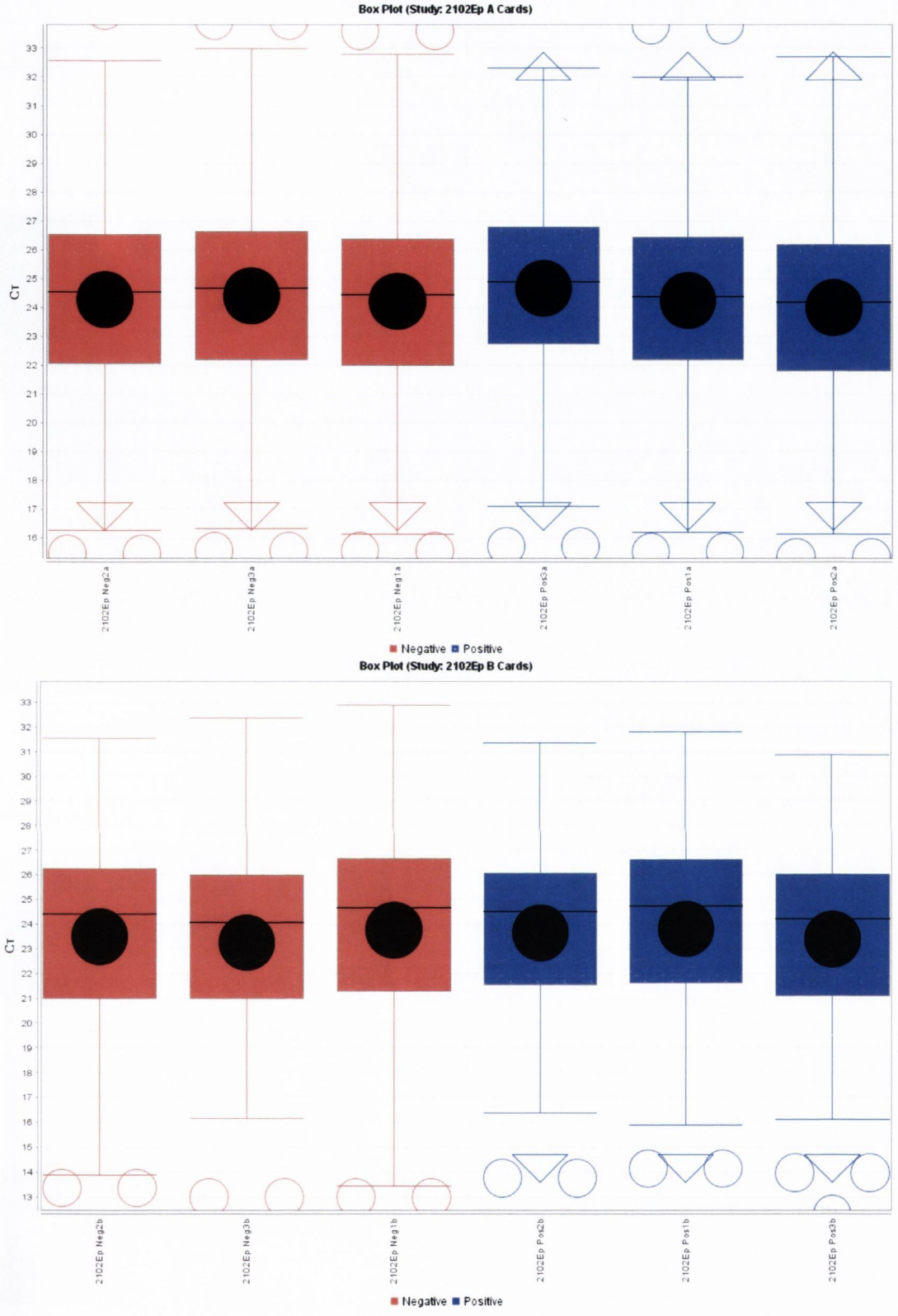


Figure 5.25—Box plots of Ct distribution gamut of expressed miRNA in 2102Ep cells. Each box represents one biological replicate. The boxes denote the middle 50% of all Ct values. The black horizontal line is the median and the black dot is the mean. The whiskers indicate maximum and minimum values 1.5 x inter quartile range between the first and third quartile.

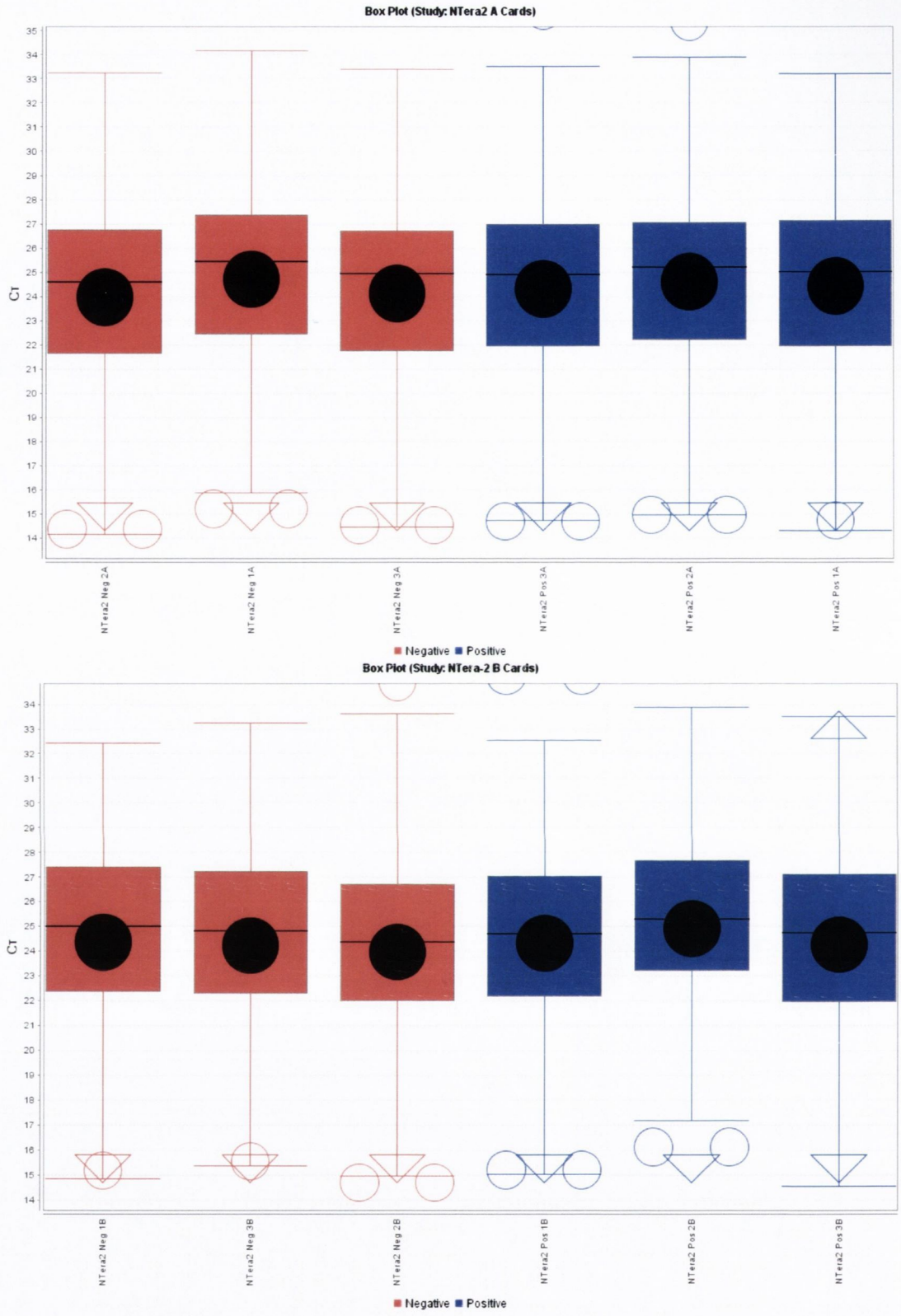


Figure 5.26—Box plots of Ct distribution gamut of expressed miRNA in NTera-2 cells. Each box represents one biological replicate. The boxes denote the middle 50% of all Ct values. The black horizontal line is the median and the black dot is the mean. The whiskers indicate maximum and minimum values 1.5 x inter quartile range between the first and third quartile.

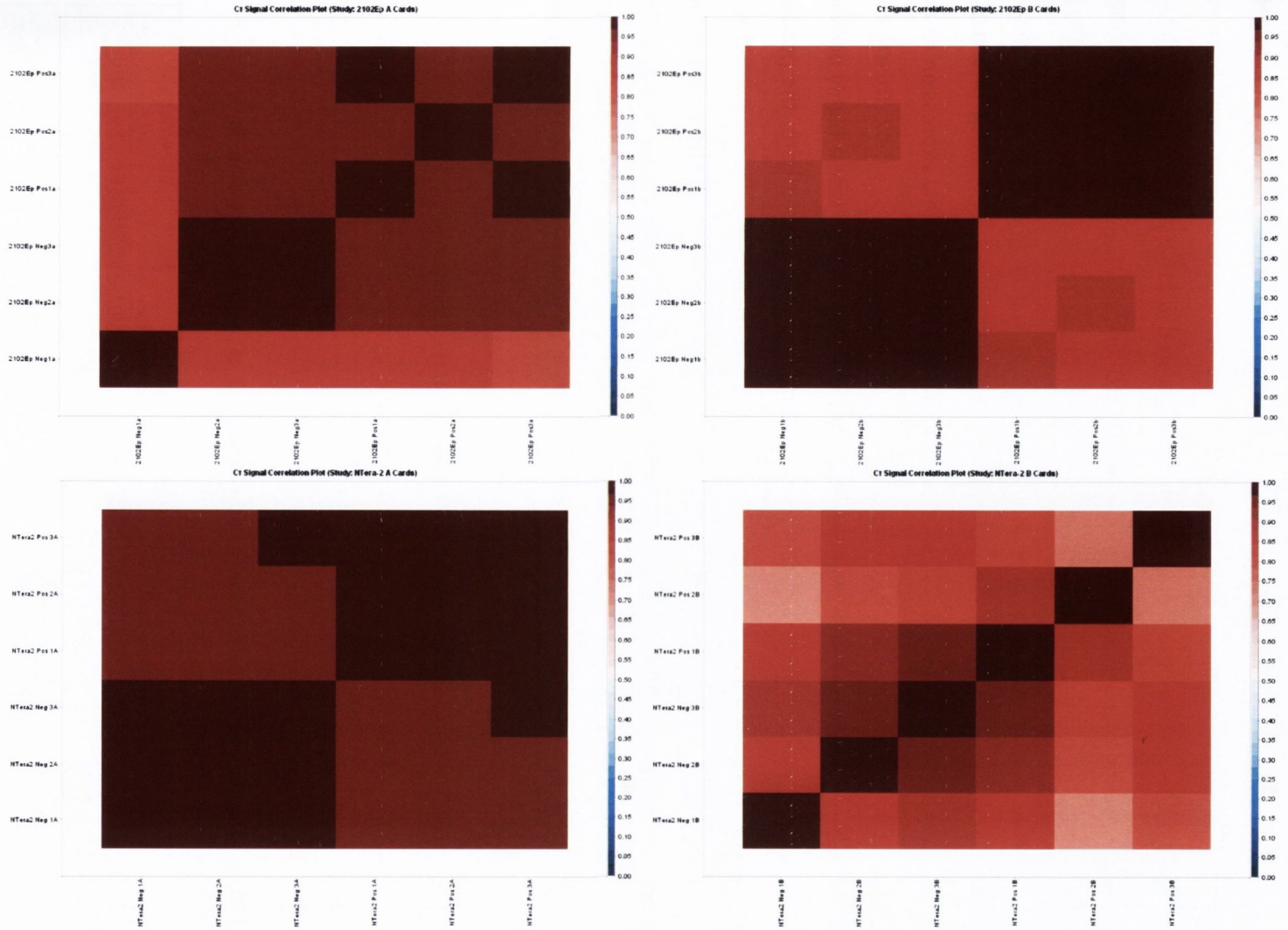


Figure 5.27—Ct signal correlation between biological replicates. The correlation of the normalised miRNA expression values is represented by these heat maps with the highest correlation depicted by dark brown.

5.4.7.3 — Analysis of Differential miRNA Expression

Differential miRNA expression was analysed with the DataAssist v3.0 software as described in Section 5.3.3.3. 2102Ep and NTera-2 cells were screened for 754 human miRNAs. The total number of miRNAs expressed and the numbers of significantly ($P < 0.05$) >2-fold up- or downregulated miRNAs in 2102Ep and NTera-2 cells are shown in Table 5.15. Total miRNA expression in both cell lines is quite similar (4% higher in NTera-2 cells than 2102Ep), however approximately twice as many miRNA were differentially regulated in 2102Ep compared to NTera-2 cells. In 2102Ep cells, the quantity of downregulated miRNA was 2.8 times higher than upregulated ones. This number was 3.3 in NTera-2 cells. These results indicate that miRNA expression is strongly biased towards downregulation upon SOX2 silencing in these cell line. In 2102Ep cells, median expression of all significantly differentially expressed miRNA ($P < 0.05$), including miRNA fold expression <2 ($\tilde{x}_{total}$), is 0.98. When only miRNA with fold expression <2 are excluded ($\tilde{x}_{>2\text{-fold}}$), the median is 0.44. In NTera-2 cells, $\tilde{x}_{total} = 0.83$ and $\tilde{x}_{>2\text{-fold}} = 0.42$. Descriptive statistics of natural-log transformed RQ data is shown in Table 5.16 and indicate a highly positive skewness statistic, a measure of the asymmetry of a distribution. As applied here, the positive skewness value indicates a significant bias of the distribution towards an downregulated fold change of overall miRNA expression. Figure 5.28 and Figure 5.29 show all >2-fold upregulated and downregulated miRNAs in 2102Ep and NTera-2 cells with their fold expression changes. The results from 2102Ep and NTera-2 cells were compared in Figure 5.30. Five unique miRNAs were significantly >2-fold differentially expressed in both cell lines, of which one (miR-181a-2*) was oppositely regulated.

Table 5.15—MiRNA expression in 2102Ep and NTera-2 cells.

	Expressed miRNAs	Upregulated	Downregulated
2102Ep	434	16	44
NTera-2	452	7	23

Table 5.16—Mean expression and data point distribution of miRNA expression in 2102Ep and NTera-2 cells.

	N	Mean	Std. Deviation	Skewness	
	Statistic	Statistic	Statistic	Statistic	Std. Error
2102Ep Total	97	-.2081	1.00116	1.148	.245
2102Ep >2-fold	60	-.3337	1.20586	1.352	.309
NTera-2 Total	62	-.4602	.77513	.947	.304
NTera-2 >2-fold	30	-.5714	1.03640	1.044	.427

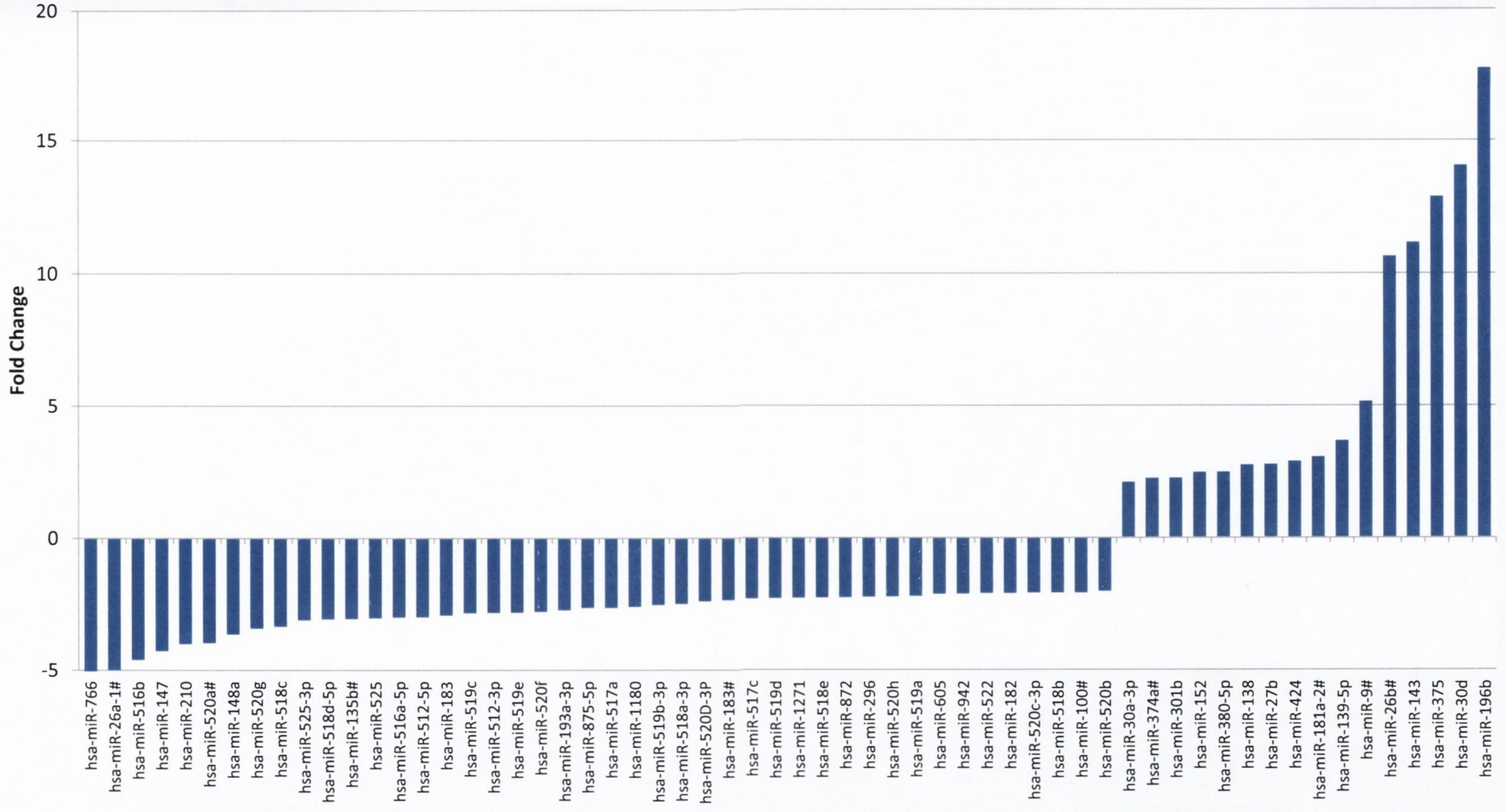


Figure 5.28—MiRNA differentially regulated in 2102Ep cells. All listed miRNAs are significantly ($P < 0.05$) >2-fold differentially expressed.

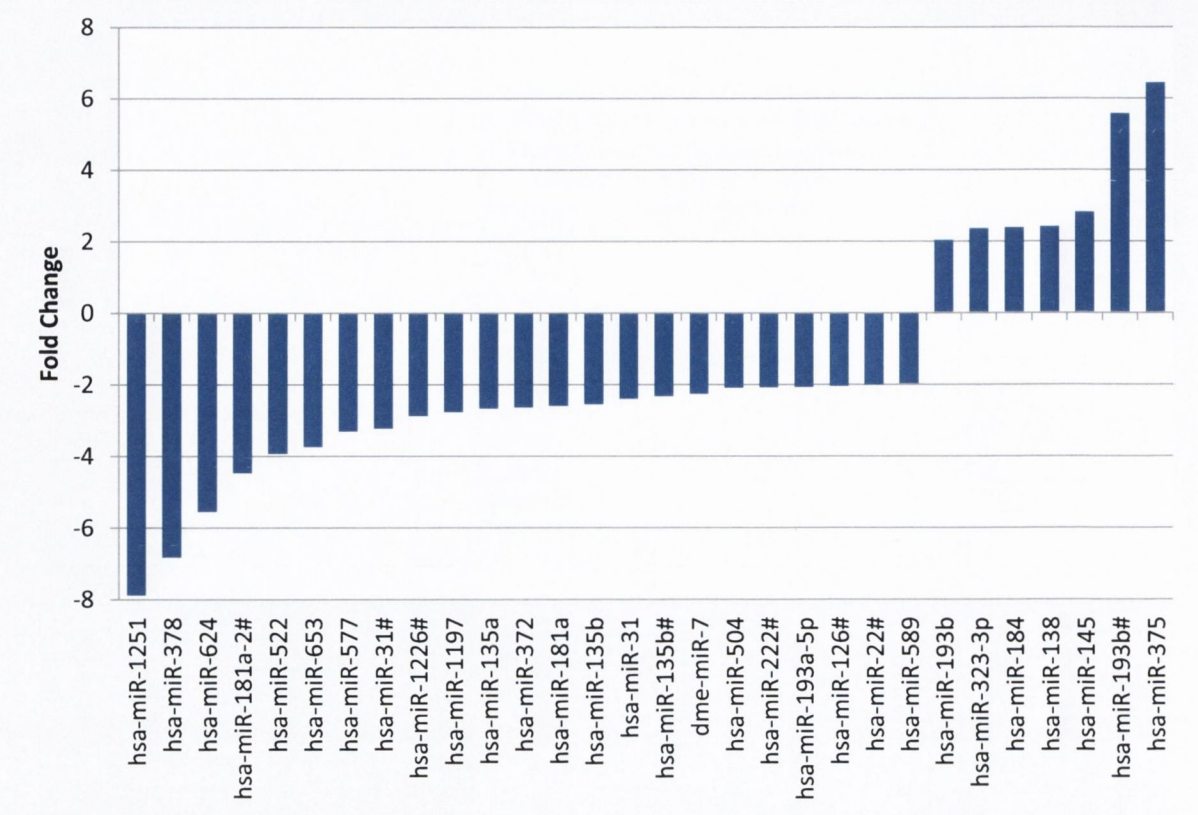


Figure 5.29—MiRNA differentially regulated in NTera-2 cells. All listed miRNAs are significantly ($P < 0.05$) >2-fold differentially expressed.

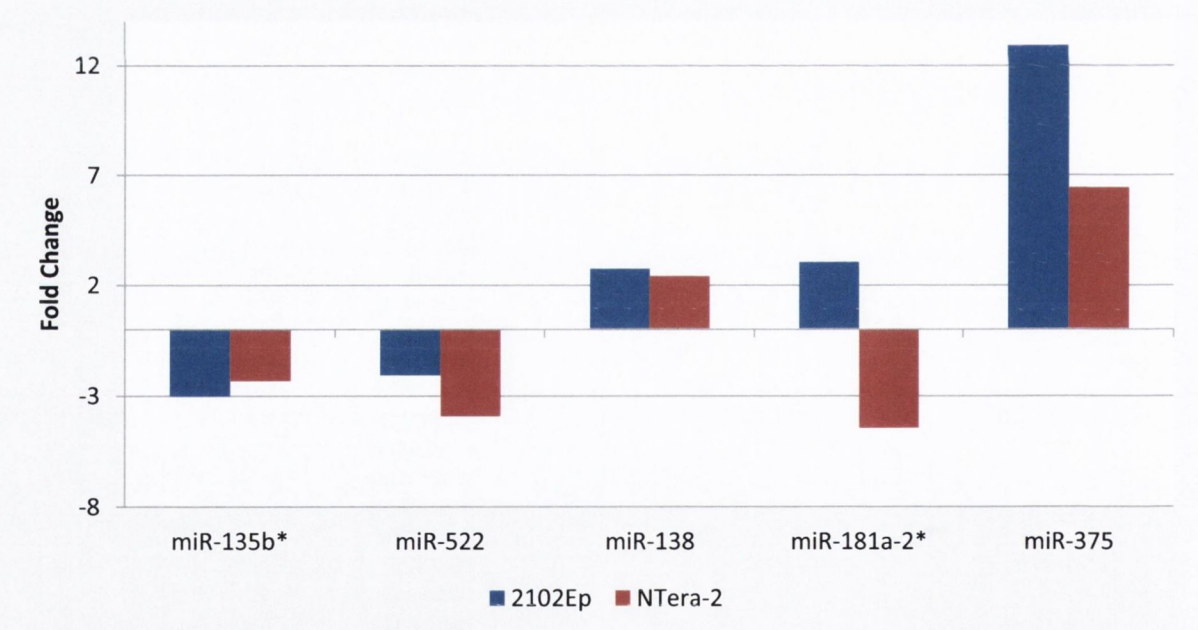


Figure 5.30—MiRNA differentially regulated in both 2102Ep and NTera-2 cells. All listed miRNAs are significantly ($P < 0.05$) >2-fold differentially expressed.

The results from differential miRNA expression in 2102Ep and NTera-2 cells upon SOX2 knock down were compared to three previous studies. The first study, by Gallagher et al. (2009), is explained in Section 3.1 and is partially comprised of a differential expression analysis of 317 miRNAs after subjecting NTera-2 cells to ATRA for three days, resulting in their differentiation of these cells and the ≥2-fold deregulation of 113 mature miRNAs.[274] The second study, by Fang et al. (2011), contains a differential expression analysis of an undefined number of miRNAs after the silencing of SOX2 in GBM cells, resulting in the ≥2-fold deregulation of 95 mature miRNAs.[466] The third study, by Marson et al. (2008), extensively mapped the *in vivo* binding sites of the core pluripotency factors, including SOX2, to defined promoter regions of miRNA in mESCs. In addition to this they also compared the expression of miRNA in undifferentiated mESCs to mouse embryonic fibroblasts (mEFs).[523] Boyer et al. (2005) mapped SOX2 binding sites in hESCs and compiled a list of intronic miRNAs that could be regulated by hES.[79] Although this study did not cover all miRNAs in the genome, some miRNAs could be regulated by the transcription of their host gene by SOX. All ≥2-fold deregulated miRNAs in 2102Ep and NTera-2 cells in this thesis, which are referenced in the positive results from the three aforementioned studies are shown in Table 5.17 and Table 5.18. The arrows (↓↑) in the 'Gallagher et al.' and 'Fang et al.' columns indicate the direction of expression of each respective miRNA in these studies. Only miRNAs bound by SOX2 in mESCs described by the Marson et al. (2008) study are compared to human analogues in Table 5.17 and Table 5.18, with the arrows indicating the direction of expression of mESCs compared to mEFs and the dot (•) indicating no change. The miRNAs found by Boyer et al. (2005) are denoted by a dot (•). These results are discussed in Section 5.5.

Table 5.17—Deregulated miRNA in 2102Ep cells which overlap with miRNA expression results from Gallagher *et al.* and Fang et al. and the ChIP results from Marson *et al.* and Boyer *et al.*[79,274,466,523]

miRNA	Fold Change	P-value	Gallagher et al.	Fang et al.	Marson et al.	Boyer et al.
hsa-miR-30d	14.06	1.09E-02		↓		
hsa-miR-143	11.18	4.94E-02		↑	•	
hsa-miR-9*	5.16	4.00E-04	↑	↓	↓	
hsa-miR-181a-2*	3.05	1.32E-02			•	
hsa-miR-424	2.87	2.80E-03	↑			
hsa-miR-27b	2.76	8.00E-04			•	
hsa-miR-152	2.45	4.25E-02			•	
hsa-miR-301b	2.24	7.00E-04				•
hsa-miR-30a-3p	2.08	3.60E-02	↑			
hsa-miR-520b	-2.00	9.10E-03	↓			
hsa-miR-518b	-2.06	1.95E-02	↓			
hsa-miR-522	-2.08	1.64E-02	↓			
hsa-miR-182	-2.08	6.00E-03			↓	
hsa-miR-942	-2.10	1.04E-02		↓		
hsa-miR-296	-2.22	2.46E-02			•	
hsa-miR-1271	-2.26	4.78E-02		↓		
hsa-miR-519d	-2.27	2.50E-03	↓			
hsa-miR-183*	-2.32	4.67E-02			↓	
hsa-miR-875-5p	-2.61	4.85E-02			•	
hsa-miR-520f	-2.76	4.40E-03	↓			
hsa-miR-519e	-2.79	1.28E-02	↓			
hsa-miR-183	-2.90	4.06E-02	↓		↓	
hsa-miR-135b*	-3.04	4.60E-03		↓	•	•
hsa-miR-518c	-3.33	1.90E-02	↓			
hsa-miR-520g	-3.41	1.40E-03	↓			
hsa-miR-148a	-3.63	4.40E-03	↑			
hsa-miR-520a*	-3.96	1.20E-02	↓			
hsa-miR-26a-1*	-4.99	5.30E-03			↓	

Table 5.18—Deregulated miRNA in NTera-2 cells that overlap with miRNA expression results from Gallagher et al. and Fang et al. and the ChIP results from Marson et al and Boyer et al.[79,274,466,523]

miRNA	Fold Change	P-value	Gallagher et al.	Fang et al.	Marson et al.	Boyer et al.
hsa-miR-145	2.81	1.91E-02	↑	↑	↑	
hsa-miR-184	2.38	6.90E-03	↑			
hsa-miR-589	-1.98	3.83E-02		↓		
hsa-miR-22*	-2.02	3.31E-02			↑	•
hsa-miR-7	-2.27	1.02E-02				•
hsa-miR-135b*	-2.33	1.89E-02		↓	•	•
hsa-miR-31	-2.41	2.30E-03			↑	
hsa-miR-135b	-2.56	1.60E-03		↓	•	•
hsa-miR-181a	-2.61	3.43E-02			•	
hsa-miR-372	-2.65	3.29E-02	↓			
hsa-miR-135a	-2.68	8.20E-03			•	
hsa-miR-31*	-3.23	3.00E-04			↑	
hsa-miR-577	-3.31	2.30E-03		↓		
hsa-miR-522	-3.93	2.30E-03	↓			

5.4.7.4 — Computational Target Prediction for Differentially Regulated miRNA

Computational target gene prediction was performed for each miRNA with a significant fold change >2 in 2102Ep or NTera-2 cells. This was performed with the TargetScanHuman software as described in Section 5.3.3.4. Target prediction generally produces results containing hundreds genes for each miRNA. As it is impossible to efficiently list all these potential targets in this thesis and this would convey a convoluted mixture of results, a different approach was taken. This approach was based upon an assumption that predicted targets which have a high frequency among similarly regulated miRNAs, have a higher chance of being genes which are actually post-transcriptionally regulated by these miRNAs *in vivo*. Similar assumptions have been made in studies before to find trends in large miRNA datasets, which could help predict potential targets.[474,524,525] Based on this, two target lists were generated with TargetScanHuman from all the miRNA significantly >2-fold downregulated and >2-fold upregulated in 2102Ep and NTera-2 cells. The targets in the lists were subsequently ranked by frequency. A histogram of these target frequencies can be found in Figure 5.31. From these, target frequencies lists were generated from a top Section of targets with the highest frequencies. To make the results comparable, a set of targets with a roughly proportional size to the whole population was taken at certain cut-off frequencies. In downregulated miRNA target set in 2102Ep cells all targets with ≥7 occurrences were included and in upregulated miRNA target set in 2102Ep cells all targets with ≥4 occurrences were included. This produced to lists of approximately 422 targets in size. These lists were named high-frequency target (HFT) lists and subsequently analysed for enriched gene ontology with the DAVID gene ontology software tool as described in Section 5.3.2.6. Functional annotation analysis is frequently performed on predicted miRNA target lists, as qualitatively analysing each target is impossible due to the large number of them and the relatively high uncertainty of *in vivo* validity.[474,526,527] Functional annotation analysis of predicted miRNA targets provides a method to find trends in general cellular function which are regulated by miRNAs.

The aim was to find commonly and differently enriched functional gene annotation clusters between the two target lists. Unfortunately, no method could be found to give target frequency a weighted value in enrichment analysis, so all genes in these lists were treated equally. Because an intrinsic functional bias among predicted targets of randomly selected miRNAs exist, a symptom which could either be caused by the TargetScan algorithm or a natural bias in miRNA target, a random control target (RCT) list was added as a comparison. This control list contained 422 of the top high-frequency targets from a list of targets generated from a randomised selection of miRNA (excluding the miRNA of interest in this study) of similar number as the significantly up or

downregulated miRNAs. The RCT list was used to compare functional annotation enrichment bias towards groups of miRNA with similar function. In Table 5.19 and Table 5.20 the ranked Group Enrichments scores for the significantly enriched functional annotation clusters are shown. The enrichment scores show a very high significant enrichment of functional annotation clusters in the high frequency target lists, whereas the RCT lists show lower significance, indicating a functional preference among high-frequency potential targets from the significantly deregulated miRNAs in 2102Ep cells.

Table 5.19 and Table 5.20 contain a summary of functional clusters with high enrichment scores of particular relevance to the expected biological change in 2102Ep cells. Instead of listing every ontological term in the cluster, duplicate terms are removed and a summarising description of the enriched terms in the cluster is given. Clusters of very similar type between the HFT and RCT lists are compared. Cluster 1 and 2 in both down and upregulated 2102Ep miRNA target lists encompass transcription factor activity and regulation, the regulation of RNA expression and DNA binding. For the target lists from miRNA downregulated in 2102Ep cells, this also includes Cluster 3, although this cluster is focussed on transcription repression, and Cluster 6 from the target lists from miRNAs upregulated in 2102Ep cells, which contains a mix of transcription activation and repression. All these terms are highly enriched in the HFT lists generated for both downregulated as upregulated miRNA in 2102Ep cells as in their companion RCT lists. However, in the HFT lists the enrichment is substantially higher compared to the RCT lists. This would indicate the main function of miRNAs in 2102Ep cells after SOX2 knock-down is the regulation of the transcriptional network. As SOX2 directly and indirectly regulates transcription and transcription factor activity, the increased enrichment of Term Cluster 1 in the HFT list could be explained as such. Term Cluster 3 in the 2102Ep upregulated HFT list indicates a miRNA-mediated alteration of post-transcriptional mRNA processing and regulation. Furthermore, this list is more enriched in embryonic developmental embryonic morphogenesis-related annotations (Cluster 4, 5 and 7), but neuron development and differentiation (Term Cluster 8) appears "under-enriched" compared to the companion RCT list. The latter might explain a bias towards excluding the upregulation of miRNA which would otherwise target genes related to these processes. Developmental and differentiation-related terms are also over-enriched in the HFT list from miRNA downregulated in 2102Ep cells (Term Cluster 5, 6 and 7) and, interestingly in contrast to the HFT lists from miRNAs upregulated in 2102Ep cells, this includes neurogenesis, perhaps suggesting a downregulation of post-transcriptional repression of genes related to this. Finally, a notable absence of enrichment of genes related to the Wnt-pathway, whereas this is present in the companion RCT list. However, no significant representation of this pathway is found in the 2102Ep upregulated miRNA target lists. Large gene lists and highly frequent predicted targets are probably needed to determine,

with even slight accuracy, functional annotation cluster enrichment which can be expressed above the "background noise" produced by false positive targets, potential bias in the computational algorithm and possible naturally occurring regulation bias of miRNAs *in vivo* (i.e. the RCT lists). The 2102Ep miRNA differential expression results produced large lists with highly frequent targets, however this is not the case in the NTera-2 data. After comparing the smaller NTera-2 HFT lists with companion RCT lists, little over-enrichment could be distinguished. The choice was therefore made to limit this analysis to the 2102Ep samples. However, microarray-TaqMan® Array cross-referencing in Section 5.4.7.5 and a more specific look at miRNA clusters and families in the discussion of this chapter (Section 5.5.2.4) address some of the potential targets of the miRNA differentially regulated in NTera-2 cells.

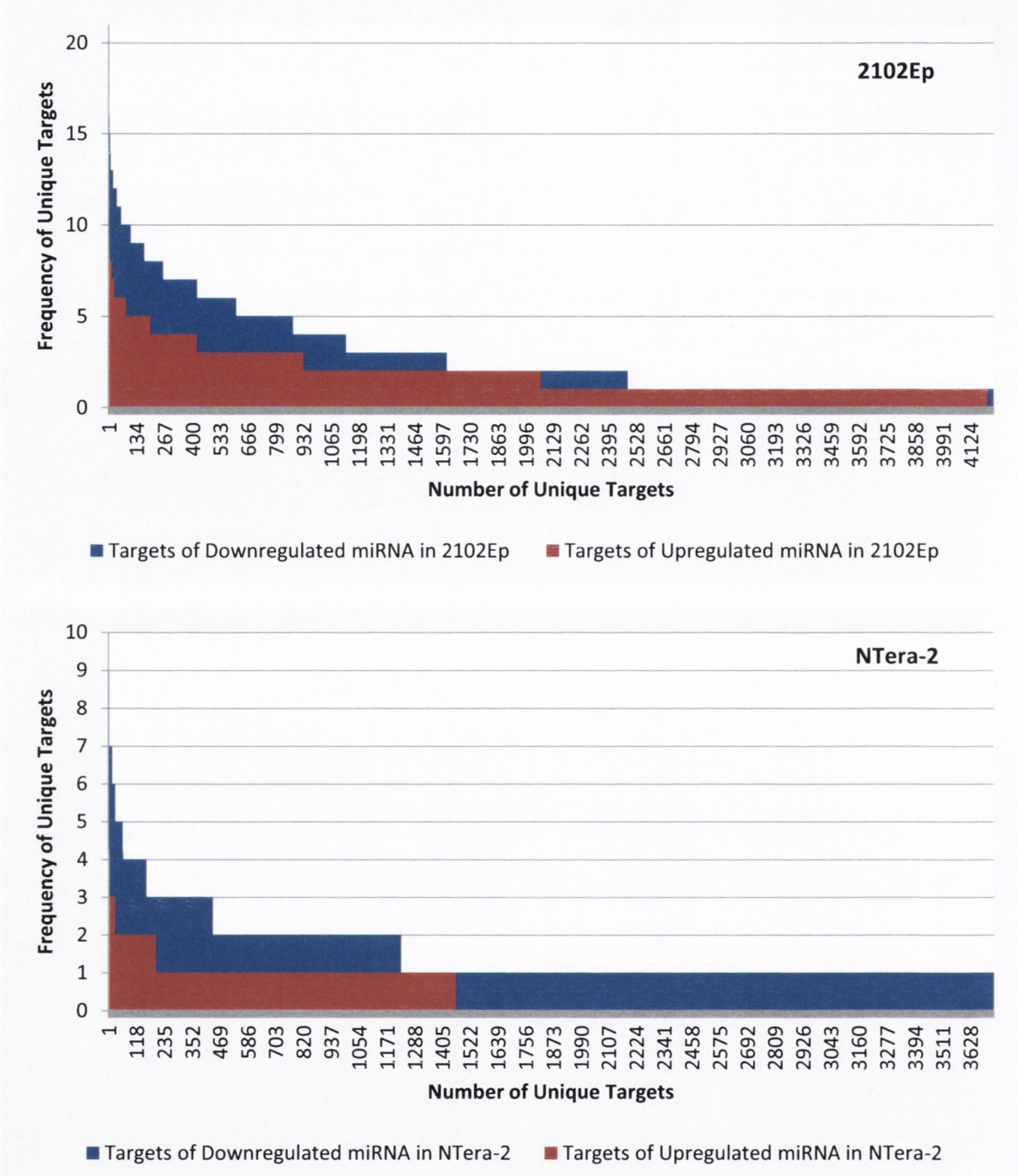


Figure 5.31—Frequency of genes as predicted targets in multiple miRNA in 2102Ep and NTera-2 cells. The X-axis represents a cumulative number of computationally predicted targets with a one or more potentially targeting miRNAs. The Y-axis represents the number of predicted miRNAs per target.

Table 5.19—Gene ontology of predicted targets of miRNA downregulated in 2102Ep cells.

Targets of miRNA downregulated in 2102Ep		Enrich. Scores	
Term Cluster	**Cluster type**	HFT	RCT
1.	Transcription, transcription regulation, DNA binding, nucleus	19.09	10.41
2.	Positive regulation of gene expression, positive regulation from RNA polymerase II promoter, positive regulation of RNA metabolic process	11.69	8.88
3.	Negative regulation of gene expression, transcription repressor activity.	5.97	3.66
4.	Zinc finger, metal binding, ion binding	4.81	2.84
5.	Lung development, respiratory development, tube development	3.88	2.74
6.	Neurogenesis, differentiation, developmental protein	2.97	1.76
7.	Gastrulation, embryonic morphogenesis, regionalisation	2.03	N/A
8.	Regulation of cell growth, regulation of cell size	1.53	N/A
9.	Wnt signalling pathway, Wnt receptor signalling pathway	N/A	2.85

Table 5.20—Gene ontology of predicted targets of miRNA upregulated in 2102Ep cells.

Targets of miRNA upregulated in 2102Ep		Enrich. Scores	
Term Cluster	**Cluster type**	HFT	RCT
1.	Transcription, transcription regulation, DNA binding, nucleus	10.26	4.21
2.	Transcription from RNA polymerase II promoter, RNA polymerase II transcription factor activity, transcription	4.61	1.83
3.	RNA-mediated gene silencing, cytoplasmic mRNA processing body	4.11	2.49
4.	*In utero* embryonic development, tube development, chordate embryonic development	3.49	0.93
5.	Embryonic morphogenesis, pattern specification process, regionalisation	3.22	1.09
6.	Transcription activator activity, transcription factor binding, transcription factor repressor, transcription factor corepressor	3.06	1.03
7.	Embryonic morphogenesis, embryonic organ development, embryonic organ morphogenesis	2.67	1.02
8.	Neuron development, neuron differentiation, neuron projection morphogenesis	2.18	3.21

5.4.7.5 — Cross-reference Analysis of mRNA and microRNA Expression Data

In Section 5.4.7.5, a group of predicted targets of the differentially regulated miRNA in 2102Ep and NTera-2 cells were ranked according to their possible relevance as inferred by the collective targeting by a large number of differentially expressed miRNA. By a similar method, in the following Section the mRNA expression data was cross-referenced with the miRNA expression data. Initially, the differentially expressed mRNAs were cross-referenced with the predicted target lists from Section 5.4.7.5, as generated by TargetScan 5.1. This produced a list of potential miRNA targets which were differentially regulated at the mRNA level. As previously described in Section 1.8, miRNA have the capability of reducing mRNA levels by RISC-induced RNA cleavage. Early genome-wide microarray studies have revealed that even single miRNAs can degrade a wide range of mRNAs simultaneously.[178] Recent studies show that, in mammals, mRNA degradation, rather than non-degradational translational suppression, is the predominant mechanism of proteinogenic inihibition by miRNAs, even accounting for ≥84% of all miRNA-mediated protein suppression.[528,529] Genome-wide mRNA expression analysis is relatively uncomplicated and inexpensive to conduct, compared to whole-proteome analysis. These attributes have resulted in an increasing number of studies which employ a combination of whole-genome expression analysis with an analysis of the differential expression of single or multiple endogenous or exogenous miRNAs.[474,475,530–534]

The general objective of an integrated miRNA and mRNA target analysis is to find *in vivo* miRNA targets with a statistical inverse correlation study of miRNAs and their respective computationally predicted targets, cross-referenced with the produced list of differentially expressed mRNAs. However, a number of challenges exist with this type of study. Due to the large number of false positive results generated by computational miRNA target prediction (i.e. noisy data), it can be difficult to find statistical correlations between miRNA expression and the expression of its targets. This problem has been highlighted before and some studies refine the data with additional criteria, such as selecting only highly altered target mRNAs for statistical correlation or including only RISC-bound mRNAs and miRNAs.[474,475,534] Another interference are secondary miRNA functions in addition to their primary function of translational suppression and mRNA destabilisation. The existence of positive correlations between miRNAs and their respective targets exist, which can be coincidental (i.e the cell's transcriptional and translational input can obviate *in vivo* miRNA regulation), but can also exist through a mechanism of miRNA-induced mRNA translation.[475] Also, as target current miRNA target prediction software does not simulate miRNA hybridisation to the coding or 5'UTR regions, a conceivably large proportion of miRNA-target pair interactions might be missed. These factors can further obfuscate the detection of

relevant miRNA-target pairs and the incomplete understanding and computational application of *in vivo* miRNA-target pair interaction requires particular strategies for a statistical approach to miRNA-mRNA correlation. Previous studies have detected subtle miRNA-target correlations by various approaches that were based on certain assumptions. The most obvious assumption is existence of a general bias in global miRNA-target pair correlation data that would be slightly negative. This global approach is difficult due to the aforementioned caveats and has been applied with mixed success on large, sometimes refined datasets.[475,532] This approach is ideally used in experiments measuring across multiple time-points, experimental conditions or specimens as it is then possible to detect the transient or per sample correlation of individual miRNAs and mRNAs in addition to the correlation of global miRNA and miRNA expression. This combinatorial approach can help detect subtle miRNA-mRNA correlations in intrinsically noisy data, but this can be prohibitively expensive due to the proportionally larger amount of consumables required and is generally applied to high-variance multi-specimen experiments, such as clinical studies.[533]

The study performed in this Section contained one experimental condition (3 days after silencing SOX2) which was compared to a control condition. The differential mRNA expression datasets from 2102Ep and NTera-2 cells were cross-referenced with the predicted miRNA target lists from Section 5.4.7.4. Table 5.21 shows the number and percentage of genes and miRNA predicted by TargetScan to be involved in targeting from the total of ≥2-fold deregulated transcripts in both cell lines. Interestingly, as a fraction of the total, fewer genes and miRNAs are predicted to be hybridise in NTera-2 cells compared to 2102Ep cells. Collectively, 657 miRNA-mRNA pairs were generated 2102Ep cells and 63 miRNA-mRNA pairs were generated for NTera-2 cells. The large discrepancy in miRNA-mRNA pairs is expected between the two cell lines is expected as $\frac{402\times60}{657}=\frac{131\times30}{x}$ $\rightarrow x\approx50$ targets, which is close to the predicted number of 63 targets. Global miRNA-mRNA expression correlation with the total miRNA and mRNA differential expression values was performed by Pearson's (parametric) and Spearman's correlation (non-parametric/ranked), but no significant negative or positive correlation was found. Recent studies found that, in particular, highly expressed mRNAs are regulated by some miRNAs.[475,529,535] Global correlation analysis was repeated with only highly expressed mRNAs, but no significant correlation was measured.

Table 5.21—The stated number of ≥2-fold deregulated genes which are predictably targeted by a stated number of ≥2-fold deregulated miRNA compared to the total number of ≥2-fold deregulated genes and miRNAs.

	Total deregulated genes	Targeted genes	Total deregulated miRNAs	Targeting miRNAs
2102Ep	402	181 (45%)	60	53 (88%)
NTera-2	131	42 (32%)	30	17 (57%)

A previous study used a target-centric approach to find statistically probably miRNA-mRNA pairs by reasoning that an increase in miRNAs targeting a single mRNA could increase the likelihood and substantiality of this mRNA's subjugation to miRNA-mediated regulation.[474] An adaptation of this target-centric approach was tested with the data used in this Section. The lists of genes described in the previous Section were ranked according to their associated number of ≥2-fold differentially regulated miRNAs. Figure 5.32 and Figure 5.33 show a set of genes ranked by the number of their associated miRNA in conformity with the aforementioned criteria. As expected, the 2102Ep data shows a substantially larger number of genes targeted by more miRNAs compared to the NTera-2 data. The assumption stating that a larger number of potentially targeting miRNA per mRNA is connected to the proportional likelihood that the expression of this mRNA is regulated by miRNA was tested by searching for an increasing trend in statistical skewness of the distribution of differential target expression values measured against the number of targeting miRNA per gene. For 2102Ep cells this trend is shown in Figure 5.34 and indicates an increase in skewness when the distributions only include targets with an increasingly higher number of associated miRNA. The 'natural' skew of a distribution of all 402 genes in the 2102Ep dataset is positive (i.e. significantly 1.499, $P < 0.05$) with the mean differential expression at around 1 fold. The mean differential expression of the genes in the groups listed in Figure 5.34 is approximately 1.5-1.9 fold, with no apparent per-group correlation. This difference could possibly be explained by a weighted bias caused by the larger amount of downregulated than upregulated miRNAs in the 2102Ep cells. However this cannot easily be statistically verified, due to high variance in the data. To summarise: in 2102Ep cells, a substantially greater number of miRNA were downregulated than upregulated, the mean of the miRNA-only target expression does not shift positively by much compared to the complete gene expression data set and the skewness in correlation with a more stringent inclusion of targets is increasingly positive (i.e. the expression values have an increasing tendency to populate the right tail of the distribution). These observations are likely to be a strong indicator that the differential expression of genes and miRNA in the 2102Ep datasets are likely to be related and that, indeed, a larger number of targeting miRNAs per gene favours this gene's regulation by miRNA, compared to those genes target by lower numbers of miRNA. The skewness of groups of genes targeted by ≥1 or ≥2 miRNAs have a lower skewness (0.984 and 1.287 respectively) than that of all 402 genes (1.499). This could be explained by the shift of average differential expression of the latter group of 1-fold to the 1.6 and 1.9-fold average differential expression of the two former groups respectively. The shift in average differential expression compensates the skewness of these two groups, resulting in a lower skewness of the overal distributions. Groups containing genes with ≥3 targeting miRNAs show increasingly larger skewness relative to the skewness of all 402 genes.

For groups limited to genes targeted by 7 miRNAs and higher, skewness is no longer statistically significant, which is probably caused by the small sample size and consequentially lower statistical power. This is also a problem for the NTera-2 data, which also contains gene and miRNA expression data which is more equally distributed between up and downregulated expression, making it more difficult to detect miRNA-induced bias. Consequently, no significant skewness of the data in a particular direction was detected.

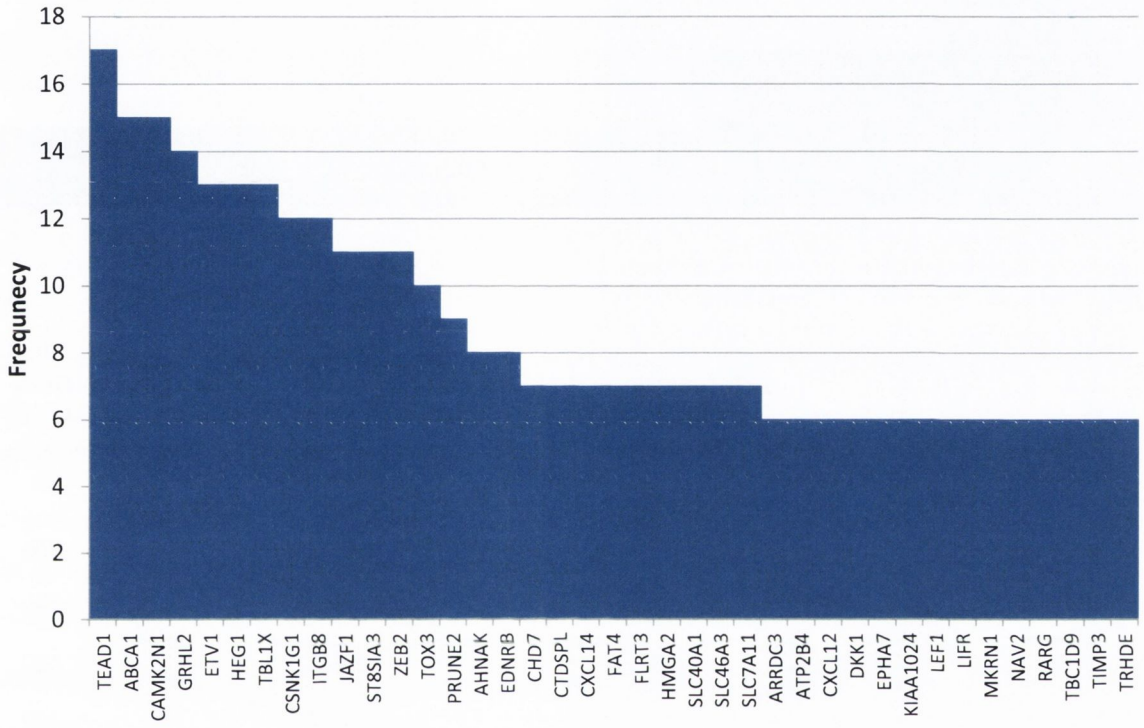


Figure 5.32—Frequency of predicted miRNA-mRNA pairs per differentially regulated gene in 2102Ep cells. Potential miRNA targets which are are also differentially regulated ≥2-fold in 2102Ep cells are displayed on the horizontal axis. The vertical axis indicates the frequency of unique ≥2-fold differentially regulated miRNA that target these genes. Only genes targeted by ≥6 miRNAs are shown.

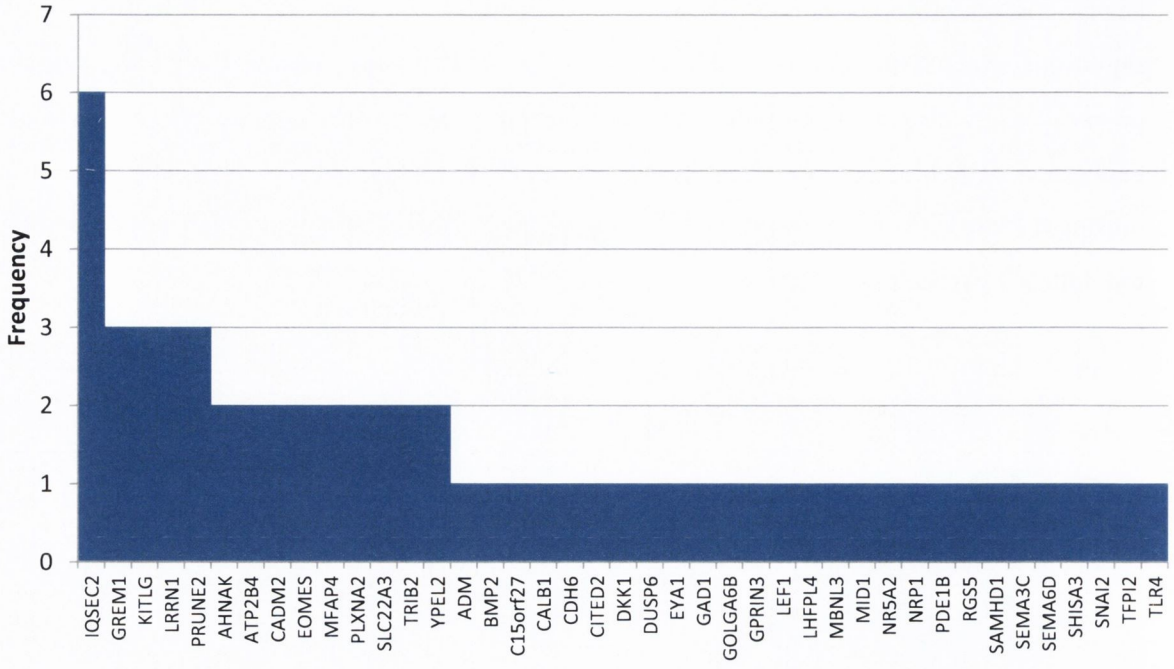


Figure 5.33—Frequency of predicted miRNA-mRNA pairs per differentially regulated gene in NTera-2 cells. Potential miRNA targets which are are also differentially regulated ≥2-fold in NTera-2 cells are displayed on the horizontal axis. The vertical axis indicates the number of ≥2-fold differentially regulated miRNA that target these genes.

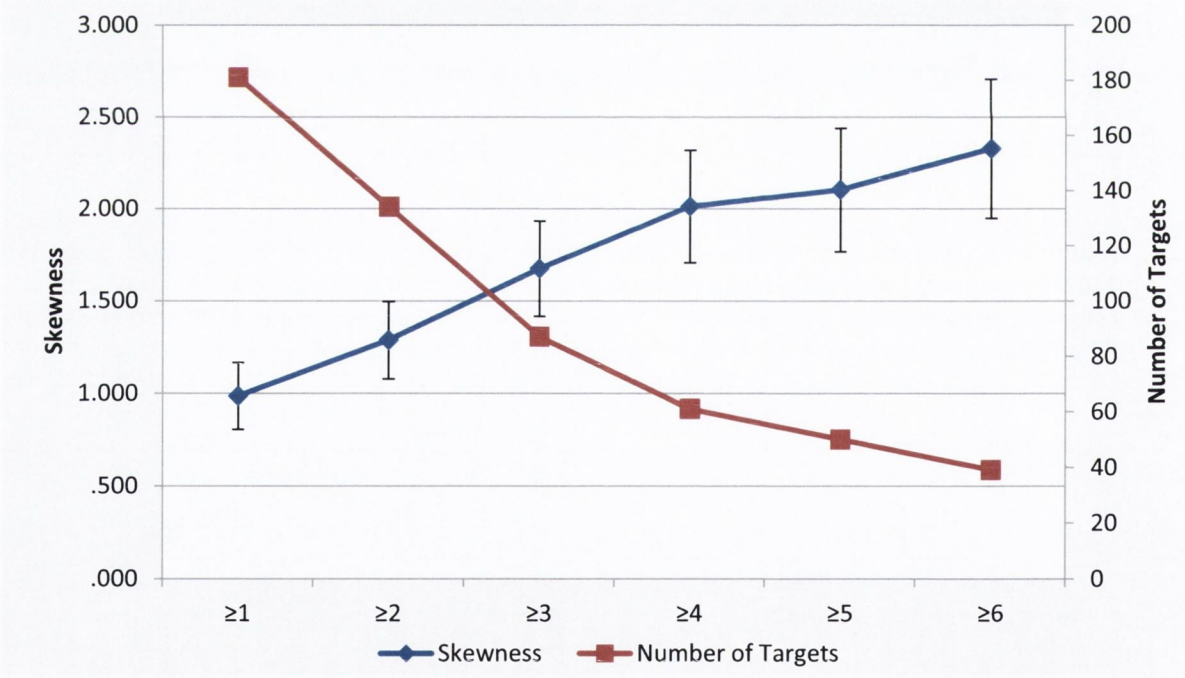


Figure 5.34—Skewness of the distributions of expression values of predicted targets in 2102Ep cells. The left vertical axis represents the degree of skewness (arbitrary value) and the right vertical axis represents the number of targets included in each group as defined by the minimum number of potentially targeting miRNA (horizontal axis). Skewness is significant for each group ($P < 0.05$). The error bars indicate the standard error of the mean.

Based on the conclusions from the previous paragraph, a statistical target-centric approach was taken to identify genes in the 2102Ep and NTera-2 gene expression data which were likely to be regulated by potentially associated miRNA found in the miRNA expression data. As an adaptation method from a previous study, the natural log-transformed expression distribution of miRNA which target a particular gene of high frequency (≥5 occurrences in 2102Ep cells and ≥ 3 occurrences in NTera-2 cells) was compared to the natural log-transformed expression distribution of all expressed miRNAs in the samples using the ranked Mann-Whitney U test to ameliorate the influence of large outliers.[474] The aim was to find differentially expressed genes that are targeted by a collection of differentially expressed miRNAs of which the median expression significantly deviated from the aforementioned control group medians. The list of genes targeted by miRNA with a collective differential expression that significantly deviates from that of the control group are found in Table 5.22 and Table 5.23for 2102Ep and NTera-2 cells respectively. From the 2102Ep datasets, which included all targets with ≥5 associated miRNAs, 30 genes were possibly significantly influenced by miRNA targeting. Of these 30 genes, 15 candidate showed an inverse gene expression vis-a-vis median miRNA expression relationship. From the NTera-2 datasets, which included all targets with ≥3 associated miRNAs, two genes were possibly significantly influenced by miRNA targeting. However, of these two genes, no candidates showed in an inverse gene expression vis-a-vis median miRNA expression relationship. This could indicate that perhaps some of these genes are indeed regulated by miRNA, while others are not. An additional explanation could be that some genes are negatively regulated by their associated miRNA, while others might be positively regulated. These results are further discussed in Section 5.5.2.3.

Table 5.22—Genes in 2102Ep cells targeted by a set of miRNA with a significantly deviating median.

Gene	Fold Expression	No. Targeting miRNAs	Median miRNA Expression	Mean miRNA Expression	Significance of Deviation from Control Group
ABCA1	-2.40	15	-2.27	-1.80	P = 0.01
CAMK2N1	2.76	15	-2.27	-1.73	P < 0.0001
GRHL2	2.60	14	-2.24	-1.36	P = 0.003
ETV1	-2.22	13	-2.27	-1.89	P < 0.0001
HEG1	2.81	13	-2.32	-1.95	P = 0.001
JAZF1	-2.02	11	-2.27	-2.08	P < 0.0001
ST8SIA3	-2.45	11	-2.27	-1.85	P = 0.002
TOX3	-2.49	10	-2.14	-1.27	P = 0.013
PRUNE2	-2.24	9	-2.14	-1.27	P = 0.013
EDNRB	-2.34	8	-2.28	-1.81	P = 0.003
CXCL14	3.00	7	-2.27	-1.73	P = 0.008
FAT4	-2.40	7	-2.27	-1.89	P = 0.008
FLRT3	2.43	7	-2.69	-2.01	P = 0.005
SLC40A1	2.51	7	-2.27	-2.42	P < 0.0001
SLC46A3	2.15	7	-2.27	-2.53	P < 0.0001
SLC7A11	-2.68	7	2.46	2.84	P = 0.002
ATP2B4	2.32	6	-2.83	-2.26	P = 0.01
CXCL12	-2.14	6	-2.28	-2.45	P < 0.0001
DKK1	19.34	6	-2.28	-2.45	P < 0.0001
LEF1	13.03	6	-2.57	-2.86	P < 0.0001
LIFR	2.91	6	5.42	5.26	P < 0.0001
MKRN1	-2.24	6	-2.32	-2.50	P < 0.0001
TBC1D9	2.28	6	-2.24	-1.79	P = 0.022
TRHDE	-2.19	6	-2.57	-2.61	P < 0.0001
BAMBI	2.57	5	-2.27	-2.45	P = 0.001
FGD4	-2.27	5	-2.27	-2.45	P < 0.001
ITGA5	3.53	5	2.46	2.22	P = 0.032
MMP2	3.88	5	-2.27	-2.48	P = 0.001
NRP1	2.52	5	2.25	1.77	P = 0.036
TBX3	5.64	5	-2.27	-2.48	P = 0.001

Table 5.23—Genes in NTera-2 cells targeted by a set of miRNA with a significantly deviating median.

Gene	Fold Expression	No. Targeting miRNAs	Median miRNA Expression	Significance of Deviation from miRNA Control Group
IQSEC2	-2.24	6	-2.64	P < 0.0001
PRUNE2	-2.12	3	-2.56	P = 0.011

5.5 — Discussion

As a core pluripotency factor in ESCs and ECCs and master regulator of embryonic and adult neural stem cells, SOX2 is crucially implicated in stem cell maintenance, among other less defined functions. In hESC cells, ChIP analysis has shown that more than 1200 genes are bound by SOX2 and can thus be regulated by this transcription factor.[79] The combinatorial miRNA and gene network that is regulated by SOX2 has not yet been characterised in ECCs.

In this study, siRNA-mediated silencing of SOX2 in 2102Ep and NTera-2 cells was followed by a 3-day incubation after which whole-genome and miRNAome expression analysis was compared to cells which were transfected with a ssNC#1 non-targeting control and co-incubated on the same plate. This chapter had three main aims. The first aim was to successfully profile the SOX2-regulated and differentiation-related transcriptome and miRNAome in 2102Ep and NTera-2 cells. The second aim was to analyse and compare the expression profiles of the two cell lines so potentially novel SOX2-linked mechanisms related to CSC and ESC biology could be identified. Finally, the third aim was to cross-reference the differential gene and miRNA expression data to potentially identify novel high-value and probably miRNA targets in 2102Ep and NTera-2 cells.

5.5.1 — Whole-genome Differential Expression

5.5.1.1 — Core Pluripotency Markers

The whole-genome differential expression profiling of 2102Ep and NTera-2 cells showed both similarities and differences in the way core pluripotency markers respond to SOX2 downregulation. One of these, NANOG, was significantly downregulated in both cell lines by approximately 2-fold according to the arrays. Analysis of these genes by qRT-PCR revealed a higher downregulation of NANOG by 2.8-fold in 2102Ep cells and 3.8-fold in NTera-2 cells. In both cell lines some NANOG pseudogenes were also significantly downregulated, but despite being close, they are not differentially expressed beyond the 2-fold threshold (data not shown). It is uncertain whether these genes and their degree of differential expression have any biological relevance to 2102EP or NTera-2 differentiation. Surprisingly, OCT4 had a much weaker response to SOX2 knock-down as it and some of its pseudogenes only show a significant expression shift of 1.3 to 1.4-fold in both cell lines (data not shown). The strong downregulation of NANOG by silencing SOX2 was expected as these genes directly coregulate each other's expression.[77–80] OCT4 appears to have a delayed reaction beyond three days to SOX2 silencing in ECCs, although it is very rapidly downregulated by the same process in NCCIT cells.[378] Compared to the pluripotency marker expression results in Section 4.4.2.2, which showed a significant downregulation of OCT4 in 2102Ep cells (49%) upon SOX2 knockdown, the expression data in this chapter conflicts. Despite this, NANOG and SOX2 expression were similarly reduced in both experiments. No solid explanation can be given for this divergence, but as three days into the differentiation program appears to be the time point upon which OCT4 rapidly downregulates, as observed in this laboratory and others, it could be possible that slight external factors could transiently shift this temporal trigger.[375,536] It is not clear whether the same discrepancy exists at the protein level as only OCT4 mRNA expression was measured.

5.5.1.2 — A Comparison of 2102Ep and NTera-2 Cell Differentiation by SOX2 Silencing

Both cell lines share 58 differentially regulated genes which, with the exception of four genes, are differentially regulated in a highly correlated manner. This could indicate a commonly shared core set of genes that are linked closely to SOX2 and other members of the core pluripotency circuitry. Indeed, of these 58 genes, 21 are also deregulated upon the knock down of SOX2 in NCCIT cells.[378] Furthermore, of these 58 genes, 19 have been found in hESCs to be bound by one or more of the three core pluripotency markers.[79] Many genes simultaneously deregulated in both in 2102Ep and NTera-2 cells have before been characterised with important functions in embryonic development. ESCs and other cells from the early-stage embryo. Previously characterised ESC

markers or genes with a functional presence in undifferentiated ESCs which were downregulated in both 2102Ep and NTera-2 cells are NANOG, NRF3, CALB1, TLR4, PHC1, NR5A2, HRASLS3, PIPOX and PRDM14.[335,481,520,537–544] Jointly upregulated genes that were previously characterised in early differentiating ESCs and lineage commitment are LGR5, HAND1, EOMES, APLNR, MIXL1, SNAI2, BMP2, MBNL3, CER1 and CFC1.[79,488,490,545–548] Furthermore, DKK1, CER1, HAND1 and EOMES have also been established as markers of Epi stem cells (EpiSC), an immediate, non-lineage differentiated derivative of ESCs.[487,543] These findings support the existence of an ESC differentiation phenotype in both 2102Ep and NTera-2 cells upon the induced downregulation of SOX2.

Despite the phenotypic similarities between differentiating 2102Ep and NTera-2 cells, obvious differences between the two cell lines also exist. Pathway analysis in Section 5.4.6.6 reveals a large representation of embryonic developmental pathways in 2102Ep cells, compared to a relative underrepresentation of these in NTera-2 cells. This seems to suggest that 2102Ep differentiate in a more ESC-like manner than NTera-2 cells. The differential expression data of both cell lines contains important markers of undifferentiated ESCs and early ESC differentiation, but many are exclusive to 2102Ep cells and few are unique to NTera-2 cells. Undifferentiated ESC markers exclusive to the 2102Ep dataset include DPPA5, KLF4, TERF1, TOX3, EDNRB, SCGB3A2, TDGF1, TFCP2L1, UTF1, CD24, USP44, AASS and ITGA6.[335,543,544,549,550] Genes frequently upregulated during early ESC differentiation towards the dermal layers which are exclusive to the 2102Ep dataset include GATA2, GATA6, NRP2, MAP2, H19, T Brachyury, AMOT, WNT5A, WNT8A, BMP4, CDX2, ID2, ITGAV, ITGA5, TGFB1, CRABP2, FN1, GLI3, LHX1, FBN2, HOXB6 and TGFBR3.[335,458,488,549–556] CHD7 was uniquely upregulated in 2102Ep cells and is frequently expressed in both undifferentiated and differentiating ESCs.[543,557] CHD7 is a transcription factor that closely cooperates with the core pluripotency factors in both undifferentiated and early differentiating ESCs.[557] It functions as a non-critical 'transcriptional rheostat' and modulates gene expression in undifferentiated ESCs and is upregulated during early ESC differentiation.[557]

Genes expressed in undifferentiated ESCs which are uniquely downregulated in NTera-2 cells are LECT1 and SFRP2.[335,558] Besides the genes shared with 2102Ep cells, some uniquely upregulated genes in NTera-2 were identified that were typical markers of early ESC differentiation. These are EPHB3 and GABRA3.[559,560] Interestingly, E-cadherin (CDH1), is upregulated also upon SOX2 downregulation of NTera-2 cells. This gene is very frequently associated with ESCs and pluripotency, but also has important functions in later stages of embryonic development and its downregulation is associated with metastatic CSC development.[101,561] A feature shared by both cell lines is the upregulation of various pivotal genes for EMT. This observation has been

highlighted in Section 5.4.6.7 for NTera-2 cells. In 2102Ep cells, SNAI1/Snail, SNAI2/Slug, BMP2, LEF1, EOMES, CDH11, TGFB1 and TGFBR3, all inducers of EMT are upregulated. In NTera-2 cells, these genes, except the latter three, are also upregulated. However, contradictorily, CDH1/E-cadherin is also upregulated in this cell line. The downregulation of this gene by many of the EMT genes previously stated is a hallmark of EMT. A possible explanation could be that these are events distributed heterogeneously in the cell culture. Additionally, the significant upregulation of a substantial number of pro-EMT genes in 2102Ep cells, with the complete absence of E-cadherin deregulation might be a contributory factor to the differentiation resistance of these cells.

A striking difference between 2102Ep and NTera-2 cells is the deregulation of various HMG box genes in 2102Ep cells by SOX2 knock-down and the complete absence of this in NTera-2 cells. SOX2, itself a member of the HMG box family, is highly involved in the regulation of HMG box genes. As described in Section 5.4.6.7, HMG box genes are the highest enriched functional group among downregulated genes in 2102Ep cells and, as described in this Section, they are strongly associated with CSC biology. It appears that in NTera-2 cells, SOX2 has decoupled from this regulatory network. Whether this is a foundation of difference in oncogenic potential between the two cell lines might warrant further investigation.

Greber et al. (2007) also showed that the inhibition of the FGF signalling pathway contributed to the differentiation of NCCIT, 2102Ep and NTera-2 cells.[378] The inhibition of FGF signalling can also be found in 2102Ep cells upon the silencing of SOX2, as FGFR3 was downregulated. Although some important components of the FGF pathway, such as FGF2, FGF4, FGFR2, SPRY2 were downregulated just under the ≥2-fold threshold value in both 2102Ep cells (data not shown), the lack of response of this pathway to SOX2 downregulation is surprising, but indicates that early SOX2-induced differentiation of 2102Ep and NTera-2 cells is relatively independent from the FGF pathway.

Despite some overlap, NTera-2 cells differ quite substantially from 2102Ep cells in their response to SOX2 knock-down. A substantial number of genes that are upregulated in NTera-2 cells upon SOX2 silencing are characteristic of embryogenesis beyond gastrulation and towards later stage of tissue specification. Some of these are FOXO4, MBNL3, BOC, CITED2, SPARCL1, EYA1, FABP3, GRHL3, LGI1, SEMA3C, SLIT2.[562–571] Another indicator of more specialised differentiation of NTera-2 cells is upregulation of Gremlin (GREM1). As an antagonist of BMP signalling, this gene is normally rapidly downregulated during early stage ESC differentiation.[552] However, during limb and organ development, Gremlin is upregulated to selectively facilitate this process.[572]

Although differential gene analysis in NTera-2 cells reveals the deregulation of many indicators of pluripotency, early ESC differentiation and the ESC phenotype, a large proportion of deregulated genes also points towards later stage embryonic development and embryonic multipotency. In previous studies, NTera-2 have been described as pluripotent progenitors, with characteristics that resemble a partial commitment towards neural differentiation.[317,318,573,574] In the context of this, the enrichment of the functional annotation cluster, 'Cell/Neuron Projection Morphogenesis and Neuronal Differentiation' (see Section 5.4.6.7), as the only enriched functional annotation cluster among downregulated genes in NTera-2, stands out because it appears to provide evidence for the suggestion that NTera-2 cells could be moving away from a phenotype partially similar to CNS progenitors.

This hypothesis is further supported by the expression of two highly characterised genes, NRP1 and NTS, both of which are upregulated in 2102Ep cells and downregulated in NTera-2 cells. NRP1 is an important pan-neural and CNS progenitor and stem cell marker and is bound by SOX2 in GBM cells.[466,575,576] NTS/neurotensin is widely distributed in the developing and adult CNS and has been suggested to function as a neurotransmitter or neuromodultor.[577] In 2102Ep cells, the upregulation of NRP1 and NTS suggests that these genes are indicative of cell specialisation from a more undifferentiated precursor as mediated by the downregulation of pluripotency factor SOX2. However, the opposite effect in NTera-2 cells appears to suggest that these cells are deviating away from a CNS precursor phenotype, in which SOX2 might have more neural stem-like roles. Furthermore, SFRP2, a Wnt antagonist, is upregulated during and can strongly induce the neural commitment and differentiation of ESCs.[578] This gene is significantly downregulated in NTera-2 cells upon the silencing of SOX2. Together, these findings might suggest that the downregulation of SOX2 might interfere with the either the initiation or progression of neural differentiation of NTera-2. However, it has to be noted that DKK1, a Wnt antagonist and inducer of the neural cell fate, is highly upregulated during 2102Ep and NTera-2 cell differentiation, as measured in this chapter, but also during hESC differentiation, as previously determined.[579,580] The aspects of DKK1 expression are further discussed in Chapter 6, but is quite indicative of the complexity of SOX2's functions in NTera-2 cells. Further functional investigation would therefore be required to elucidate this mechanism and test the hypothesis mentioned at the start of this paragraph.

5.5.1.3 — Directly SOX2-linked Genes

One of the main aims of this chapter was to discover potentially novel members of the transcriptional network linked to SOX2 in embryonal carcinoma and to highlight differences of this network in 2102Ep and NTera-2 cells. Some previous studies have mapped the binding sites of

SOX2 in hES and GBM cells.[79,466] Compared with these studies, results in Section 5.4.6.4 showed that approximately 12% and 23% of differentially expressed genes in 2102Ep cells were bound by SOX2 in hES and GBM cells respectively, whereas for deregulated genes in NTera-2 cells this percentage increased to 21% and 32% respectively. When the results from the two studies were merged both these percentages increased to 31% and 46% for 2102Ep and NTera-2 cells, whereas 47% of genes deregulated in both 2102Ep and NTera-2 cells are bound by SOX2 in either hES or GBM cells. However, only 4% and 5% of genes in the 2102Ep and NTera-2 datasets, respectively, showed SOX2 binding in both hES and GBM cells. Only 20% of identified SOX2 binding sites identified in hESCs (from a total of 1247 hits) are also found in those identified in GBM cells (total hits: 3203). These results suggest that selective SOX2 binding is highly dependent on the cell type. A possible hypothesis could also state that, taking into account the proportions of the data from the hES and GBM studies, and also the very low overlap of SOX2 binding in hESc and GBM cells, when cross-referenced with the differential gene expression data in 2102Ep and NTera-2 cells, that 2102Ep cells, but more so, NTera-2 cells are in a SOX2 transcpriptional transition state between hESCs and GBM cells (neural cells). For example, if NTera-2 cells were more similar to hESCs than GBM cells, a far greater adjusted proportion of differentially regulated genes would be expected to be bound in hESCs by SOX2 compared to those bound in GBM cells. As all three studies (Boyer et al., Fang et al. and this one) all use different methods and because this thesis does not contain a ChIP-seq study of 2102Ep and NTera-2 cells, a further investigation would therefore be needed to test this hypothesis.

Table 5.24 shows all differentially regulated genes in 2102Ep and NTera-2 cells which conversely have upstream SOX2 binding sites in hES and GBM cells as defined by Boyer et al. (2005) and Fang et al. (2011) respectively. Some genes which have previously been functionally defined as direct transcriptional targets of SOX2 are shown in bold and referenced with their related publication. These are confirmed targets and do not include genes which have been tentatively predicted by high throughput statistical means. Although quite a large proportion of these genes have been postulated as transcriptional targets of SOX2, few have been validated as such. DUSP6, another possibly SOX2 regulated gene, is upregulated in SOX2-silenced 2102Ep cells and downregulated in SOX2-silenced NTera-2 cells. This gene is an inhibitor of the FGF/ERK pathway and has been found to be expressed in several tissue types.[581] However, it also appears to have crucial functions during CNS development and in the adult brain.[581,582] Longer term studies could elucidate and alterations or deviations of cell fates of NTera-2 cells after SOX2 downregulation.

Table 5.24—Deregulated genes in 2102Ep and NTera-2 cells which have SOX2 binding sites in their promoters as defined by Boyer et al. and Fang et al.[79,466]

	Boyer et al. (hESCs) Only	Fang et al. (GBM cells) Only	hES + GBM
2102Ep Only	ANP32A CCDC92 DPYSL3 FAM115A FAM124A FRAT2 GAD2 HELLS **HESX1**[583] ID1 JUP KIAA0101 MED12 RAB25 RGS4 SEPT6 **SOX2**[79] TDGF1 TRIML2 USP44 ZEB2 ZSWIM3 ZYG11A	ACOXL ADAMTS9 AFF2 AMOT ANKRD44 ATP12A BMPER BNIP3 C1orf61 C7orf46 CACNA2D3 CACNG4 CAMK2N1 CASC1 CD53 CDCA7L CDH10 CDH11 CHI3L1 CNKSR2 CTDSPL CYP24A1 DPPA5 ELMO1 EMP1 ENC1 EPHA7 ETV1 FBP1 FGD4 FRRS1 GRIA1 ID2 JAZF1 KDM4C KIAA1024 LBH LHX1 MAP2 MSX2 NAV2 NEXN NRP2 PREX1 PSTPIP2 PTPRZ1 RASGRP3 RSPO3 RYR3 SKAP2 SLC16A10 SLC23A2 SLC44A2 SP8 SPOCK1 STAB2 TBL1X TBX3 TEAD1 TFCP2L1 TMOD1 TRHDE **UTF1**[584]	ARRDC3 BAMBI BTK **CDH2**[585] FZD2 FZD7 **GLI3**[586] GRHL2 HAS2 **HOXA1**[587] PYGM SLC7A11 **TCF7L1**[499]
NTera-2 Only	C7orf16 FEZ1 KITLG LGI1 LRAT RHD RXRG SESN3 SFRP2 SLCO5A1 TRIM22	ANGPT1 BOC C18orf26 CITED2 DSCAM EPHB3 GPR141 GPRIN3 GRHL3 NFATC2 PKDCC PNMA2 POSTN SEMA3C SEMA6D SPRY1 TFPI2 TRIB2 YPEL2	CDH6 LRRN1 ST6GAL2
2102Ep + NTera-2	CER1 **DKK1**[588] DUSP6 HAND1 IQSEC2 **NANOG**[79] PIPOX **PRDM14**[589] SHC3 TIMP4	AHNAK APLNR ATP2B4 EDIL3 **NR5A2**[460] NRP1 NTF3 PLA2G16 PLXNA2 TLR4 VSNL1 WLS ZNF483	BMP2 **EOMES**[590] GAD1 NFE2L3

5.5.2 — Differentially Expressed miRNAs in 2102Ep and NTera-2 Cells

One of the main aims of this chapter was to map the direct and indirect regulatory response of miRNA expression upon induced downregulation of SOX2 in 2102Ep and NTera-2 cells. Whole-microRNAome expression analysis was performed with both 2102Ep and NTera-2 cells after SOX2 knock-down as described in Section 5.4.7. The results from this showed that the total number of expressed miRNAs in 2102Ep and NTera-2 cells were similar, with 434 expressed in 2102Ep cells and 452 expressed in NTera-2 cells out of a total of 754 miRNAs represented in the analysis. Of these, 60 and 30 miRNAs were differentially regulated in 2102Ep and NTera-2 cells respectively, a notable difference in number that, as with the whole-genome expression results, appears to suggest that SOX2 has fewer regulatory functions in NTera-2 cells compared to 2102Ep cells. However, in contrast to the differential gene expression in 2102Ep and NTera-2 cells, SOX2 appears to primarily have a function as a transcriptional activator of miRNA, as substantially more miRNA are downregulated than upregulated upon SOX2 silencing. This also reflects the findings from a previous group who found this when silencing SOX2 in GBM cells.[466] Five unique miRNAs were significantly deregulated in both 2102Ep and NTera-2 cells, specifying a relatively low overlap in miRNA function between the cell lines. Results in Section 5.4.7.3 showed that one of the overlapping miRNAs (miR-181a*) was oppositely regulated in both cell lines, which could indicate that this miRNA is regulated by SOX2 in a cell type-specific manner. All miRNAs functionally analysed in Chapter 3 were detected in 2102Ep and NTera-2 cells, except miR-10a. However, only miR-184 was found to be significantly >2-fold deregulated, which was in NTera-2 cells. Somewhat unexpectedly, miR-184 was upregulated upon SOX2 downregulation in NTera-2 cells. Some other miRNAs, including miR-184 in 2102Ep cells, were significantly downregulated, albeit below the 2-fold threshold.

Some deregulated miRNAs in 2102Ep and NTera-2 cells are intronic, i.e. their respective loci are positioned within the introns of protein-coding genes. In total, 17 intronic miRNAs are deregulated in 2102Ep cells and four of these correspond with differentially expressed host genes from the gene array results (miR-9*:C1orf62, miR-26a*:CTDSPL, miR-196b:HOXA9 and miR-766:SEPT6). 15 deregulated intronic miRNAs are found in the NTera-2 data, but none of the host genes could be found among in the differential gene expression data of this cell line.

The 2102Ep and NTera-2 differential miRNA expression results from this study were compared to a miRNA expression study by Gallagher et al. (2009), who differentiated NTera-2 cells with ATRA for three days. Of all deregulated miRNAs in 2102Ep cells, 23% are also deregulated in ATRA-induced differentiating NTera-2 cells as Table 5.17 in Section 5.4.7.3 infers.[274] Only one miRNA is oppositely regulated between these groups. Remarkably, this correlation is much higher between

these groups than when SOX2 silencing-induced differentiating NTera-2 cells are compared to ATRA-induced differentiating NTera-2 cells, which show only 13% overlap. This phenomenon also contrasts with the gene expression results, which show that both NTera-2 groups have a higher correlation (see Table 5.13). The discrepancy could be explained by the inclusion of a cluster of miRNAs belonging to the C19MC polycistron, some of which are detected and downregulated in differential NTera-2 expression results from the study by Gallagher et al. (2009). In this thesis, the miRNAs from this cluster, except one, were only detected in 2102Ep cells. A previous study emphasised the extremely low expression of C19MC miRNAs in NTera-2 cells, suggesting that these miRNAs were below the sensitivity threshold of TaqMan Arrays, while the methods adopted by Gallagher et al. (2009) were able to detect these.[591] When the group of C19MC miRNAs are ignored, differentiating 2102Ep miRNA results in this study actually have lower overlap with Gallagher et al.'s data (10%) than do the NTera-2 miRNA results.

5.5.2.1 — MicroRNAs in Cancer Stem Cells

Due to the significant overlap of normal stem cell gene and miRNA expression and that of CSCs, the differentially regulated miRNAs in 2102Ep and NTera-2 cells could have important functions in CSCs and cancer biology. A huge abundance of miRNAs have been associated with cancer, either as tumour suppressor miRNAs, oncomirs or both The following paragraph will focus exclusively on deregulated miRNAs in 2102Ep or NTera-2 cells which have defined or suggested roles in CSCs. These miRNAs are shown in Table 5.25. This table shows that, particularly in 2102Ep cells, several miRNAs with CSC function are deregulated. The regulation of previously established, CSC-related tumour suppressor miRNAs were expected to be upregulated during 2102Ep and NTera-2 cell differentiation, while similar oncomirs were expected to be downregulated. As seen in Table 5.25, this is not always the case, possibly denoting alternative malignant or tumour suppressive functions in the two ECC lines screened in this study. No deregulated miRNAs in 2102Ep or NTera-2 cells from this chapter have yet been functionally linked to these cell lines. The selection of miRNAs in Table 5.25 therefore provide interesting candidates for functional analysis in ECCs. This is further detailed in later Sections of this discussion.

Table 5.25—MiRNAs deregulated in 2102Ep and NTera-2 cells with previously documented functions in CSCs.

Cell Line	miRNA	Fold Change	Function	CSC origin
2102Ep	miR-9*	5.16	Tumour suppressor	Glioblastoma[270]
	miR-181a-2*	3.05	Oncomir	Hepatocellular carcinoma[271]
	miR-380-5p	2.46	Oncomir	Neuroblastoma[592]
	miR-301b	2.24	Tumour suppressor	Breast cancer[593]
	miR-30a-3p	2.08	Tumour suppressor	Breast cancer[594]
	miR-182	-2.08	Tumour suppressor	Breast cancer[254]
	miR-183*	-2.32	Tumour suppressor	Breast cancer[254]
	miR-183	-2.90		Pancreatic cancer[255]
	miR-135b*	-3.04	Oncomir	Colorectal carcinoma[273]
NTera-2	miR-7	-2.27	Tumour suppressor	Glioblastoma[595]
	miR-181a	-2.61	Oncomir	Hepatocellular carcinoma[271]
			Tumour suppressor	Glioblastoma[596]
	miR-372	-2.65	Oncomir	Testicular germ cell carcinoma[597]

5.5.2.2 — MicroRNAs in Cancer and Embryonic Stem Cell Function

As described in Sections 1.1 to 1.3, CSCs maintenance and proliferation often involves molecular mechanisms that are also active in undifferentiated ESCs, more committed multipotent stem cells and adult stem cells. The importance of SOX2 in ESCs has been well documented and also has frequently been associated with malignant nature of CSCs, including embryonal carcinoma and those found in other germ cell tumours.[282,405,598–601] Because of the similarities between ESCs and embryonal carcinoma stem cells, like 2102Ep and NTera-2 cell lines, a certain overlap of miRNA expression between ESCs and the miRNA expression data in this chapter was expected. Several studies have profiled enriched miRNA expression in undifferentiated ES or have compared miRNA expression of undifferentiated ESCs to more differentiated, closely descendant cell types. This has produced a characteristic miRNA signature for ESCs, containing highly expressed miRNAs and ones which only upregulate shortly after differentiation. The miRNA expression results from SOX2-silenced 2102Ep and NTera-2 cells were compared to the results produced from these studies. These miRNA signatures and positive cross-references are shown in Table 5.26 and Table 5.27. In addition, any experimentally validated functional roles of these miRNAs in cancer has been added to highlight possibly undiscovered relationships between ES and CSC biology as tied by specific miRNAs in these cancers. As expected, most miRNAs in Table 5.26 and Table 5.27 have many oncogenic or tumour suppressive functions in different cancers, signifying the overlap of molecular processes found in both ESCs as in malignancy. Several CSC-associated miRNAs from Table 5.25 (miR-9*, miR-181a-2*, miR-181a, miR-7 miR-301b) have, besides their roles in CSCs, additional oncogenic or tumour suppressive functions shown in Table 5.26 and Table 5.27 in the previous Section. However, many miRNAs shown in these tables have not (yet) directly been linked to CSC biology, despite some highly published functions in cancer and recorded associations with ESCs.

Table 5.26—Deregulated miRNAs in 2102Ep cells with enriched expression or documented functions in ESCs and ESC differentiation, including previously reported roles in cancer.

miRNA	Fold Change in 2102Ep	Enriched in	Tumour supp. /oncomir	Cancer type
miR-196b[602]	17.75	EBs	Oncomir	MLL-associated leukaemia[602] GBM[603]
			Tumour Supp.	Lymphoblastic leukaemia[604,605] Breast cancer[606] B-cell acute lymphoblastic leukaemia[605]
miR-375[413]	12.89	Definitive Endoderm	Oncomir	Prostate carcinoma[607] Breast cancer[608]
			Tumour Supp.	Gastric cancer[609] Hepatocellular carcinoma[610,611] Maxillary sinus squamous cell carcinoma[612] Oesophageal squamous cell carcinoma[613] Head and neck squamous cell carcinoma[614] Melanoma[615]
miR-143[616]	11.18	ESCs	Tumour Supp.	Cervical cancer[617] Bladder carcinoma[618–620] Colorectal cancer[621,622] Prostate cancer[623,624] Osteosarcoma[625]
miR-26b*[626]	10.65	ESCs		
miR-9*[413]	5.16	Definitive Endoderm	Tumour Supp.	Glioma[270,627]
miR-181a-2*[628]	3.05	Endothelial-differentiated ESCs	Oncomir	Hepatocellular carcinoma[271]
miR-380-5p[592]	2.46	ESCs	Oncomir	Neuroblastoma[592]
miR-520b[629]	-2.00	ESCs	Oncogene	Breast cancer[630]
			Tumour Supp.	Breast cancer[630,631]
miR-296[199]	-2.22	14-day Differentiated ESCs	Oncomir	Enriched in 22 different cancer cell lines[632]
			Tumour Supp.	Colorectal cancer[633] Breast cancer[633] Lung cancer[633] Gastric cancer[633] Parathyroid cancer[633] Liver cancer[633] Bile duct cancer[633]
miR-512-3p[413]	-2.80	ESCs	Tumour Supp.	Hepatocellular carcinoma[634]
miR-512-5p[413]	-2.97	ESCs	Tumour Supp.	Gastric cancer[635]
mir-210[636]	-3.99	Endothelial-differentiated ESCs	Oncomir	Clear cell renal carcinoma[637] Breast cancer[638] Lung cancer[639]
			Tumour Supp.	Lung cancer[640] Oesophagal squamous cell carcinoma[641]
miR-26a-1*[616]	-5.99	ESCs		

Table 5.27—Deregulated miRNAs in NTera-2 cells with enriched expression or documented functions in ESCs and ESC differentiation, including previously reported roles in cancer.

miRNA	Fold Change in NTera-2	Enriched in	Tumour supp. /oncomir	Cancer type
miR-375[413]	6.43	Definitive Endoderm	Oncomir	MLL-associated leukaemia[602] GBM[603]
			Tumour Supp.	Lymphoblastic leukaemia[604,605] Breast cancer[606] B-cell acute lymphoblastic leukaemia[605]
miR-145[207]	2.81	EBs	Tumour Supp.	Glioma[642] Bladder cancer[618] Malignant melanoma[643] Breast cancer[644] Colorectal cancer[645] Lung adenocarcinoma[646,647] Endometrial adenocarcinoma[648] Oesophagal squamous cell carcinoma[392] Prostate cancer[649,650]
mir-126*[651]	-2.05	ESCs		
miR-222*[616]	-2.08	ESCs		
miR-7[652]	-2.27	Neural-differentiated ESCs	Tumour Supp.	Hepatocellular carcinoma[653] Glioma[654] Non-small lung cancer[655]
miR-135b[348]	-2.56	ESCs and 7-day differentiated ESCs		Colorectal cancer[656]
mir-181a[628,657]	-2.61	Endothelial-differentiated ESCs	Oncomir	Hepatocellular carcinoma[271]
			Tumour Supp.	Oral squamous cell carcinoma[658] Lung cancer[659] Gastric cancer[659] Glioma[660]
miR-372[580,616]	-2.65	ESCs	Oncomir	Colorectal cancer[580] Gastric cancer[661] Testicular germ cell carcinoma[597]
mir-135a[657,662]	-2.68	mESCs and mEBs	Oncomir	Non-small cell lung carcinoma[663] Hodgkin lymphoma[664] Colorectal cancer[656]
miR-183[253,665]	-2.90	mESCs	Tumour Supp.	Pancreatic cancer[255]
miR-181a-2*[628]	-4.47	Endothelial-differentiated ESCs	Oncomir	Hepatocellular carcinoma[271]
miR-378[616]	-6.84	ESCs	Oncomir	Lung cancer[666] Breast cancer[667] Glioblastoma[668]

Several miRNAs with notable interactions and functions are listed in Table 5.26 and Table 5.27. Of interest are miRNAs which have either been previously known to be regulated by SOX2 or are known to have SOX2 binding sites in their promoters.

For example, miR-181a and miR-181a-2* have been previously associated with hepatocarcinoma and glioblastoma stem cells, but have a tumorigenic function in the first cancer type and a potentially tumour suppressive role in the second.[271,596] MiR-181a-2* is upregulated in SOX2-silenced 2102Ep cells, but is, together with miR-181a, downregulated in NTera-2 cells. Both miRNAs have binding sites for SOX2 in their promoters regions in mESCs, suggesting a possibility of direct, but cell-specific regulation by this transcription factor in other cell lines, such as ECCs, which could be contributory or disadvantageous to carcinogensis.[523]

MiR-7 appears to have a potentially tumour suppressive role in glioblastoma stem cells, but its expression is downregulated upon the knock-down of SOX2 in NTera-2 cells.[595] In hESCs, SOX2 is bound to the promoter region of miR-7, suggesting that SOX2 directly regulates miR-7 expression.[79] miR-7 is expressed during neural differentiation and has been shown to orchestrate important mechanisms in immature and adult neurons.[359,652,669] These findings reveal a potentially important link between SOX2 and neural development and function, but also in CNS cancers. Due to the downregulation of miR-7 after SOX2 knock-down in NTera-2 cells, these findings additionally support the hypothesis that SOX2 downregulation guides NTera-2 cells away from a committed neural fate. The downregulation the CNS-enriched miRNA, miR-135a, in NTera-2 cells, further supports this hypothesis.[665]

MiR-9*, an upregulated miRNA in SOX2-silenced 2102Ep cells, has previously been revealed to regulate SOX2 expression in glioma stem cells.[270] This miRNA is likely also regulated by SOX2 as binding sites for this transcription factor exist in its promoter region in mESCs.[523] This suggests the existence of a direct double-negative feedback loop between SOX2 and miR-9*.

MiR-135a and miR-135b are both downregulated in NTera-2 cells upon SOX2 knock-down, signifying their potential importance in these cells. In ESCs, miR-135a regulation appears to be contextual, as one study found it be to downregulated in differentiating mESCs and another study showed high enrichment of this miRNA in undifferentiated mESCs.[657,662] In addition miR-135a is also expressed in the murine brain.[665] Although miR-135a is less well studied in human development, miR-135b has been found to be enriched in a hESC line, but is also further upregulated during early differentiation of this cell line.[348] Conversely, compared to fully differentiated tissue, this miRNA is upregulated in the differentiating hESC line, suggesting a transient role of miR-135b in early differentiating hESCs, besides possible functions in

undifferentiated hESCs.[348] In 2102Ep cells, the minor strand of miR-135b, miR-135b*, is downregulated, but no published functions could be found for this miRNA. Interestingly, two studies showed that the promoter regions of miR-135a and miR-135b/135b* are bound by SOX2 in both hES and mESCs.[79,523] MiR-135a and miR-135b have oncogenic properties in various cancers, as shown in

Table 5.27, supporting the possibility that SOX2 could directly mediate carcinogenicity of miR-135a/b expressing cancers, in addition to directly controlling miR-135a/b in ESC function and differentiation.

As mentioned in the previous paragraphs, from the miRNAs in Table 5.26 and Table 5.27, miR-145 and miR-9* have been previously been shown to have the capacity to directly regulate SOX2.[207,627] Also, in previous research, the regulation of miR-145 by SOX2 has been suggested, but not confirmed.[466] This SOX2-miR-145 autoregulatory loop, as it might also exist in NTera-2 cells, could be a direct mechanism seeing that SOX2 has been found to bind miR-145 in mESCs. Furthermore, in mESCs, miR-9* has a SOX2 binding site in its promoter and is upregulated in 2102Ep cells upon SOX2 knock-down. This suggests that SOX2 acts as a repressor of miR-9*, which in turn, can repress SOX2 expression. The existence of this direct double-negative autoregulatory loop might be cell specific due to the substantial downregulation miR-9* in GBM cells upon SOX2 silencing.[466] Moreover, in NTera-2 cells, miR-9* substantially upregulates upon the introduction of ATRA, but not upon SOX2 knock-down.[601] The existence of CSCs in GBM cell populations has been widely published and are phenotypically similar to normal stem cells of the CNS.[120] Based on the previous findings and given the regulation of miR-9* in hESC-like 2102Ep and NTera-2 cells and in neural stem cell-like GBM cells, the hypothesis could be held that annotates miR-9* as a transient regulator of SOX2 during differentiation of hESCs into GBM cells. However, this cannot be confirmed without further studies.

Finally, miR-126 has been shown to have the capacity to directly inhibit post-transcriptional SOX2 expression.[670] In NTera-2 cells, only the minor strand of this miRNA, miR-126*, is ≥2-fold downregulated (2.1-fold). However, miR-126 is 1.5-fold downregulated in NTera-2 cells. If this fold change is biologically relevant, this could point to, perhaps an indirect, autoregulatory loop between SOX2 and another miRNA, a mechanism potentially worth further investigation.

5.5.2.3 — The Identification of MicroRNA Targets by Computational Analysis

The third aim of this chapter was to find possible mRNA targets of deregulated miRNAs in 2102Ep and NTera-2 cells, which were themselves deregulated by miRNA-induced degradation. As supposed, based on the results of previous studies, this mRNA degradation could be assessed by a measurable difference in mRNA expression as determined by gene array analysis.[474,475,530–534] The first step of this method was to generate lists of potential targets for each ≥2-fold differentially expressed miRNA in 2102Ep and NTetra-2 cells. For this, TargetScan 5.1 was used as this computational target prediction software was rated as one of the most reliable in a survey of multiple different target prediction programs.[330] In addition, the widespread use of TargetScan in the field of miRNA target prediction, meant that the produced results would be quite comparable with other studies.

In Section 5.4.7.4, the predicted targets from all ≥2-fold deregulated miRNA in 2102Ep or NTera-2 cells were collectively analysed for functional annotation enrichment. The group of highly represented genes were compared to a random control set. Based on this analysis, deregulated miRNAs in 2102Ep cells showed a likely enrichment of functions in regulation of transcription, embryonic development and morphogenesis, whereas, more specifically, the downregulated miRNAs were under-represented for functions in Wnt signalling compared to the control group of miRNAs (suggesting a possible miRNA 'bias' towards suppressing Wnt signalling) and the upregulated miRNAs were under-represented for functions in neuronal development and differentiation (suggesting a possible miRNA 'bias' towards reducing suppression of neuronal development). These observations match the functional annotation analyses of differential gene expression as seen in Section 5.3.2.6 and signify the central importance of miRNAs in the cellular processes concerning the cell-specific nature of 2102Ep and NTera-2 cells. Unfortunately, this method yielded insignificant results for NTera-2 cells, possibly due to the smaller size of the miRNA target dataset.

To identify potential miRNA targets among the 2102Ep and NTera-2 differential gene expression data, the assumption was made that the likelihood of miRNA regulation of a target gene was correlated with the number of *in vivo* miRNAs that potentially target this gene. In 2102Ep cells, this assumption was supported by an inverse correlation of target expression skewness and the number of miRNA-mRNA, indicating a non-random influence of miRNA-mRNA interactions on overall mRNA expression at higher miRNA-mRNA pairs per target gene. The results from the 2102Ep data suggested that almost half of all deregulated genes were potentially targeted nearly all deregulated miRNAs. These fractions were lower in NTera-2 cells, with approximately a third of all deregulated genes targeted by more than half of all deregulated miRNAs. Although no

biological explanation can be definitively given for this discrepancy, the deregulated mRNA:miRNA ratio in the NTera-2 data compared to 2102 data does indicate a lower probability of achieving a proportional number of positive matches. With the method outlined in Section 5.4.7.5, lists were compiled containing genes with a high possibility of being regulated by miRNA in 2102Ep or NTera-2 cells after SOX2-silencing. In 2102Ep cells, 30 target genes were significantly highlighted statistically, of which 13 were inversely correlated with the median expression of their affiliated miRNA. Although the results for these 30 target genes reject the null-hypothesis, which states that the expression values of targeting miRNAs are randomly distributed in the total group, the median expression of 15 are actually positively correlated with their collective targets. This could indicate that, although the used method might be applicable to predict some miRNA-mRNA interactions, these interactions must be more qualitatively assessed. Depending on many known and unknown variable, some miRNAs probably have more functional influence on gene expression than others. This more qualitative approach has been suggested and, till an extent, applied in previous research.[474,475]

The 15 target genes in 2102Ep cells with an inverse relationship with their collective of targeting miRNAs are listed in Table 5.28. Among these 15 genes, SLC7A11 has been validated as a target of miR-30a, denoting a high probability of also being regulated by miR-30d.[535] This list of genes provides candidates for further validation as targets of their respective miRNAs. Among these are some previously discussed genes with important roles in stem cell function, differentiation and embryonic development (e.g. DKK1, MMP2 and BAMBI).

Besides the potential miRNA targets, listed in Table 5.28, a number of differentially expressed genes in 2102Ep and NTera-2 cells are validated targets of differentially expressed miRNAs. Any positive results that could be found in the large body of miRNA target research are listed in Table 5.29. Upregulated miRNAs and genes in 2102Ep or NTera-2 cells are coloured green and those which are downregulated are coloured red. Table 5.29 show many inverse miRNA-target associations, but also several positively correlated miRNAs and targets. Of the latter, miR-145 has previously been validated as a positive regulator of MYOCD.[671]

Table 5.28—Inversely regulated target genes and their associated miRNAs.

Gene	Fold Change	Median miRNA Expression.	Targeting miRNAs
CAMK2N1	2.76	-2.27	*miR-1271, miR-182, miR-183*, miR-30a-3p, miR-30d, miR-519a, miR-519b-3p, miR-519d, miR-520a*, miR-520b, miR-520c-3p, miR-520d-3p, miR-520f, miR-520g, miR-520h*
GRHL2	2.60	-2.24	*miR-1271, miR-143, miR-182, miR-30a-3p, miR-30d, miR-512-3p, miR-519d, miR-520a*, miR-520b, miR-520c-3p, miR-520d-3p, miR-520f, miR-520g, miR-520h*
HEG1	2.81	-2.32	*miR-1271, miR-143, miR-182, miR-30a-3p, miR-30d, miR-512-3p, miR-519d, miR-520a*, miR-520b, miR-520c-3p, miR-520d-3p, miR-520f, miR-520g, miR-520h*
CXCL14	3.00	-2.27	*miR-520a*, miR-520f, miR-520d-3p, miR-519d, miR-520c-3p, miR-520b, miR-9**
FLRT3	2.43	-2.69	*miR-183, miR-193a-3p, miR-27b, miR-512-5p, miR-520g, miR-520h, miR-522*
SLC40A1	2.51	-2.27	*miR-519a, miR-519b-3p, miR-519d, miR-520a*, miR-520b, miR-520c-3p, miR-520d-3p*
SLC46A3	2.15	-2.27	*miR-519d, miR-520a*, miR-520b, miR-520c-3p, miR-520d-3p, miR-520g, miR-520h*
SLC7A11	-2.68	2.46	*miR-143, miR-148a, miR-152, miR-27b, miR-30a-3p, miR-30d, miR-380-5p*
ATP2B4	2.32	-2.83	*miR-1271, miR-148a, miR-152, miR-183, miR-26a-1*, miR-520f, miR-380-5p*
DKK1	19.34	-2.28	*miR-519a, miR-519b-3p, miR-520a*, miR-520b, miR-520c-3p, miR-520d-3p*
LEF1	13.03	-2.57	*miR-512-3p, miR-520a*, miR-520b, miR-520c-3p, miR-520d-3p, miR-766*
TBC1D9	2.28	-2.24	*miR-424, miR-519a, miR-519b-3p, miR-519d, miR-520g, miR-520h*
BAMBI	2.57	-2.27	*miR-519d, miR-520a*, miR-520b, miR-520c-3p, miR-520d-3p*
MMP2	3.88	-2.27	*miR-519a, miR-519b-3p, miR-519d, miR-520g, miR-520h*
TBX3	5.64	-2.27	*miR-519a, miR-519b-3p, miR-519d, miR-520g, miR-520h*

Table 5.29—Previously validated targets of miRNAs deregulated in 2102Ep and NTera-2 cells. Green miRNAs/genes were upregulated. Red miRNAs/genes were downregulated.

Cell Line	miRNA	Validated Target Genes
2102Ep	miR-145	KLF4, MYOCD, HOXA9, BNIP3[671–673]
	miR-182	RARG[674]
	miR-196b	HMGA2[675]
	miR-210	HOXA1, HOXA9, KIAA1161[676,677]
	miR-26a/*	MAP2, HMGA2, BMP2[678–680]
	miR-296	NANOG[199]
	miR-30a	HMGA2, SLC7A11, LHX1[535,679,681]
	miR-30d	LHX1[681]
	miR-9*	SOX2[627]
	miR-520h	ID1, ID3[682]
NTera-2	miR-184	NFATC2[683]
	miR-31	DKK1[684]
	miR-181a	VSNL1[685]

5.5.2.4 — Functional Relevance of MicroRNA Clusters

MiRNA clusters and families are highly conserved regulatory elements in the genome and function and are regulated in an orchestrated manner. A large number of deregulated miRNAs in 2102Ep and NTera-2 cells belong to particular miRNA clusters and families. MiRNA clusters are defined as collection of miRNA loci positioned in relative proximity in the genome and frequently co-regulated in similar fashion by a single promoter region. The term miRNA cluster is quite loosely defined. If these miRNAs are expressed from a single transcript (pri-miRNA), their cluster is often referred to as a polycistron. However, deeper investigation can sometimes reveal additional separate promoter regions for individual miRNAs within a cluster, resulting in separately transcribed miRNAs, which can be even located on overlapping pri-miRNAs.[686] A miRNA family consists of a group of highly similar or homologous miRNAs which have genetically evolved from a common ancestor and share a common seed sequence. The gene loci of these miRNAs can either reside within a cluster or can be located elsewhere in the genome, even on different chromosomes. They often have highly comparable functionality, but can be regulated differently.

The differentially expressed miRNAs in 2102Ep and NTera-2 cells were grouped into clusters and families, which are listed in Table 5.28. From this table it is immediately obvious that particular miRNA families and clusters are highly co-regulated and it is likely that they have important functionality to cellular changes as a consequence of SOX2 silencing. Table 5.30 contains miRNA clusters and families with two or more deregulated mature miRNAs in 2102Ep and NTera-2 cells. The red coloured miRNAs are downregulated in the cell lines, whereas the green coloured miRNAs are upregulated. The miRNAs listed in between brackets are miRNAs that are <2-fold significantly downregulated in 2102Ep or NTera-2 cells. Although these miRNAs lie below the 2-fold threshold, they still might be functionally significant due to their potential contribution towards the entire family or cluster. Furthermore, a large majority of these miRNAs are deregulated <2<1.5-fold.

Table 5.30—Deregulated miRNA families and clusters in 2102Ep and NTera-2 cells. Green miRNAs were upregulated. Red miRNAs were downregulated.

2102Ep	C19MC cluster	*miR-512-3p, miR-512-5p, miR-516a-5p, miR-516b, miR-517a, miR-517c, miR-518a-3p, miR-518b, miR-518c, miR-518-5p, miR-518e, miR-519a, miR-519b-3p, miR-519c, miR-519d, miR-519e, miR-520a*, miR-520c-3p, miR-520d-3p, miR-520f, miR-520g, miR-520h, miR-522, miR-525, miR-525-3p (miR-517b, miR-518f, miR-520b)*
	miR-183-96-182 cluster	*miR-182, miR-183, miR-183* (miR-96)*
	miR-26 family	*miR-26a-1*, miR-26b* (miR-26a)*
	miR-30 family	*miR-30a-3p, miR-30d (miR-30c)*
	miR-374a-545 cluster	*miR-374a*, (miR-545*)*
	miR-148/152 family	*miR-148a, miR-152*
	14q32.31 cluster	*miR-380-5p, (miR-370, miR-1197, miR-539)*
	miR-302/372/373/ 520be/520acd-3p family	*miR-520a*, miR-520b, miR-520c-3p, miR-520d-3p (miR-302c, miR-372)*
NTera-2	C19MC cluster	*miR-522 (miR-517c)*
	miR-7-1197-3529 cluster	*miR-7, miR-1197*
	miR-135 family	*miR-135a, miR-135b, miR-135b**
	miR-193 family	*miR-193a-5p, miR-193b, miR-193b**
	miR-221/222 family + cluster	*miR-222* (miR-221, miR-222)*
	14q32.31 cluster	*miR-323-3p, miR-1197 (miR-487a)*

The categorising of some deregulated miRNAs in clusters and families constructs a functional map of miRNA deregulation as induced by SOX2 silencing in 2102Ep and NTera-2 cells. The regulation of whole miRNA families and clusters signify functional importance to cell processes after SOX2 silencing, as the functional roles of the members of these families and clusters are often related. The dominant miRNA cluster in 2102Ep cells, as can be seen in Table 5.30, is the polycistronic C19MC cluster, which maps to the 19q13.41 genomic region and is the largest miRNA cluster in the mammalian genome.[686] This cluster, which consists of 46 tandemly repeated miRNAs, has been found to be enriched in hESCs, and in human placenta and foetal brain tissue, but its biological importance to these has yet to be fully defined.[175,687–690] Interestingly, the C19MC cluster is frequently amplified or activated in various cancers, including embryonal tumours, and the oncogenic potential of some of its members has been previously confirmed.[691–693] 28 members of the C19MC cluster are deregulated in 2102Ep cells, all of which are downregulated. In 2102Ep cells, the deregulated miRNAs from this cluster potentially target 82 deregulated genes in 2102Ep cells, of which 51 are upregulated, as predicted by TargetScan 5.1. This represents more than one fifth of all deregulated genes in 2102Ep cells, but the ratio of upregulated genes to downregulated genes in the group of candidates that are potentially targeted by the deregulated miRNAs from the C19MC cluster is similar to the ratio of the entire population. Thus, from the gene array data, no accurate conclusions can probably be drawn about how the C19MC cluster regulates gene expression in 2102Ep cells and further target validation would be required to

determine which transcripts are repressed by this cluster and which, if any, might be activated. Despite this, among the 82 deregulated genes, particular candidates of biological relevance could be initially selected for functional validation. Functional annotation analysis of the 82 potential targets with DAVID shows the three highest enriched terms to be tissue morphogenesis (P = 0.00017; FRAS1, KLF4, TBX3, BMP2, HAND1, LEF1, TCF7L1, TFAP2A, TGFBR3, ZEB2), Wnt receptor signalling pathway (P = 0.0015; CSNK1G1, DKK1, FZD7, LEF1, RARG, TCF7L1, TBL1X, WNT5A) and mesenchymal cell development (P = 0.0014; BMP2, EDNRB, LEF1, NRP2, TGFBR3, ZEB2). Almost all of the genes related to these terms—as shown in brackets—and other annotations of high significance are upregulated upon SOX2 silencing of 2102Ep cells, which correlates with the downregulation of the C19MC miRNAs and therefore might be good initial candidates for target validation. Furthermore, the downregulation of this cluster upon SOX2 silencing and its expression in the foetal brain both suggest that SOX2 could be a regulator of C19MC, but a characterisation of this cluster's promoter region(s) and further SOX2 binding studies would be required to confirm this. The C19MC cluster contains various miRNA families, some of which are related to miRNAs elsewhere in the genome. The miR-520 miRNAs in the C19MC cluster belong to the larger miR-302/372/373/520be/520acd-3p family, of which some other members are differentially regulated in 2102Ep cells. Cross-functionality of these miRNAs have previously been suggested and linked to hESC differentiation.[629] A member of this cluster, miR-302c, has been previously been shown to be regulated by the OCT4/SOX2 dimer and downregulates upon hESC differentiation.[266] The significant upregulation of this miRNA in 2102Ep cells is unexpected, but as this is only slight (1.3-fold), especially compared to the differential regulation of the other members of the miR-302/372/373/520be/520acd-3p family, the effect of this could perhaps be biologically irrelevant. All members of the miR-302 cluster are expressed in both 2102Ep and NTera-2 cells, but, apart from miR-302c, remain unchanged. In differentiating NTera-2 cells, the miR-302 cluster only begins to downregulate after approximately 8 days after the induction of differentiation by ATRA.[694] The lack of miR-302 downregulation by SOX2 silencing is supported by previous research which showed that miR-302 downregulated by OCT4 or OCT4/SOX2 silencing, but not by SOX2 silencing alone.[266] Finally, only two members of the C19MC cluster (miR-522 and miR-517c) are differentially expressed, and downregulated, in NTera-2 cells. A previous analysis showed that the expression of the C19MC miRNAs was very low in this cell line, compared to hESCs and almost no differential expression was detected upon the differentiation of NTera-2 cells.[591] The apparent lack of C19MC miRNAs in NTera-2 cells is confirmed by the data in this thesis (not shown) and is also reported by previous research conducted in this laboratory, which showed that the majority of C19MC miRNAs did not deregulate during NTera-2 cell differentiation or were not detectable by qRT-PCR.[274] Previous research and the findings in this chapter strongly

suggest that the role of the C19MC cluster is more similar between 2102Ep cells and hESCs than NTera-2 cells.

Other differentially regulated clusters and families in 2102Ep and NTera-2 cells notable. The miR-183-96-182 cluster has been implicated with cancer and CSCs and its associated miRNAs are downregulated in 2102Ep cells after SOX2 knock-down. In 2102Ep cells, the miRNAs in this cluster potentially target 48 deregulated transcripts in 2102Ep cells. Although it is overexpressed in some malignancies and has a proposed oncogenic role, it has also been found to be highly underexpressed in some CSC types, including the ECC line, Tera-2, from which NTera-2 cells were originally derived.[254,695,696] Members of the miR-183-96-182 cluster could not be detected in NTera-2 cells according to the data used for this chapter (not shown). The reason of this discrepancy between 2102Ep cells and NTera-2 cells and other CSCs might warrant further investigation due to the apparent selectivity of miR-183-96-182 function in various cancer types and cell lines, including a possible diagnostic and therapeutic value from this. On chromosome 14, a large miRNA bipartite cluster of 7 and 46 miRNAs exists in the genomic 14q32.31 region.[697] Some members from this region were deregulated in both 2102Ep and NTera-2 cells after SOX2 knock-down. The deregulated miRNAs from this region were all upregulated in 2102Ep cells, while mixed differential expression of these was mixed in NTera-2 cells. However, no 14q32.31 region miRNAs were commonly regulated in 2102Ep and NTera-2 cells and, likewise, no deregulated genes in both cell lines were common potential targets of the combined set of miRNAs. The 14q32.31 miRNA cluster has been previously been proposed to be the largest tumour suppressor miRNA cluster in the genome and its suppression has been associated with various cancer types, but also with the adult and embryonic stem cell state, making it an interesting candidate target for future CSC research.[424,697,698]

This chapter comprehensively charts the gene and miRNA expression networks linked to 2102Ep and NTera-2 cells, while comparing the similarities of, but also many differences between these networks. This has shown that differential gene and miRNA expression in SOX-silenced 2102Ep cells is more similar to the equal response of hESCs, than when the latter are compared to NTera-2 cells. While differential gene and miRNA expression in 2102Ep cells resembles early differentiation in the three germ layers, NTera-2 cells deviate away from their neuronal commitment upon the knock-down of SOX2.

A set of genes and miRNAs with SOX2 binding sites in their promoter regions, but with no previous transcriptional association with this transcription factor, has been identified. Among differentially expressed miRNAs, some have previously been shown to regulate SOX2 protein expression. The relationships between SOX2 and these miRNAs warrant further functional

investigation due to the potential central nature of this network in stem cell potency, embryonic development and some cancers.

A statistical and a qualitative approach to target prediction was taken by combining differential gene and miRNA expression, with which a list of high-probability targets was compiled. These candidate targets can be used for future functional validation, among which genes and miRNAs with important roles in cancer, CSCs and pluripotency would be of direct relevance to this project.

In summary:

- The SOX2 knock down-induced differentiation of 2102Ep and NTera-2 cells resulted in the deregulation of many genes important to cancers stemness, particularly those of the Wnt and TGFβ pathways.
- This also deregulated oncogenic and tumour suppressive miRNAs that are predicted to regulate many of the differentially expressed genes.
- Many SOX2-targeting miRNAs were also deregulated, indicating the activity of SOX2-miRNA autoregulatory loops.
- The differentiation of 2102Ep and NTera-2 cells, as induced by SOX2-silencing, had characteristics of hESC differentiation. However, this was more closely resembled by 2102Ep cells, while NTera-2 cells appeared to differentiate away from a committed neuroectodermal phenotype.

Chapter Six

GENERAL DISCUSSION

6— General Discussion

Current cancer theory holds that the origin, proliferation, metastasis, therapy resistance and recurrence of malignant neoplasms are fundamentally engendered by an often relatively small subpopulation of neoplastic cells which exhibit a phenotype enriched with stem cell-like characteristics. Despite extensive research, controversy remains about the origin and the conception of stem-like cancer cells. Additionally, the fraction of these cells that contribute to the growth and proliferation of tumours have been shown to substantially differ among various cancer types.[157,163,164] The relatively new field of CSC research, as kicked off by the discovery of leukaemia CSCs, has brought about these questions and others, while greatly expanding the understanding of the function and contribution of stemness in malignancy.[53] The specific targeting of CSCs in tumours is a tough challenge due to their adaptability and resistance to different therapies, their ability to survive and thrive in otherwise harsh environments and their ability to evade detection. Despite the challenges, CSC-targeted therapy is certainly achievable. As with conventional cancer therapy, a big obstacle to surmount is being able to selectively target the tumourigenic component for treatment—in this cased CSCs—while avoiding unacceptable side-effects and damage to normal cells. The difficulty of CSC-selective targeting, besides the aforementioned caveats, is the similarity of the CSC-phenotype to that of normal stem cells.

Despite the difficulties, a few pioneering studies performed proved that the targeting of CSCs was possible, while leaving normal stem cell populations unaffected.[429,699] These studies showed that the drug-induced upregulation of particular tumour suppressors in LSCs led to a reduction in the proliferation of these cells, while leaving their normal counterparts, haemopoeitic stem cells, unaffected by the same conditions. As with gene-centric CSC-specific therapy, the therapeutic targeting of miRNAs in CSCs is difficult. The expression profiles of miRNAs in normal stem cells and their malignant doppelgangers are almost always very similar, possibly because of their relatively broad scope of post-transcriptional regulation, which more broadly influence processes important to both cell types. Although miRNAs with exclusive functions in CSCs, compared to NSCs, have yet to be found, several miRNAs have recently been characterised to have crucial functions in cancer, CSCs and NSCs. As mentioned in Section 1.12, at the time this study began, a small number of miRNAs were identified as oncomirs or tumour suppressor miRNAs with explicit CSC and NSC functions. These were the tumour suppressive let-7, miR-200 families and the oncogenic miR-302 and miR-17-92 clusters.

This dissertation focused primarily on the biomolecular and (epi)genetic fundamentals of CSCs and more specifically on some of those potentially contributing to embryonal carcinoma cells, the

stem cell component of 87% of non-seminomatous germ cell cancer.[700] The research conducted was founded on previous studies conducted by colleagues in this laboratory, who profiled differential gene and miRNA expression in NTera-2 cells during early ATRA-induced differentiation.[274] To keep results comparable to the aforementioned studies and results produced by other laboratories, the majority of experiments performed for this thesis with 2102Ep and NTera-2 cells, involved measurements made after subjecting the cells to three days of experimental conditions. This timeframe was also chosen based on previous studies which indicated that phenotypic and morphological change of ES and ECCs, including the initiation of pluripotency factor downregulation, became apparent 2-3 days after the induction of differentiation.[297,317,374,536] The focus on CSC-specific differentiation was primarily held for the gene-centric approach in this study, described in Chapter 4. No comparative studies were found or performed which could point towards CSC-specific miRNAs and the function and expression of many miRNAs deregulated during ATRA-induced differentiation of NTera-2 cells, were still broadly uncharacterised in relation to stemness or cancer stemness. Due to the highly inclusive nature of gene regulation by miRNAs it could even be possible that the existence of CSC-specific miRNAs is highly uncommon. Up till date, research has shown that stem cell-like cancer cells rather appear to hijack certain components of the miRNA machinery with crucial functions in normal stem cells. Because of these aspects, the miRNA selection process in this thesis did not necessarily exclude miRNAs that were shown to have some possible functions in normal stem cells.

6.1 — Discussion of the Findings

In Chapter 3, five miRNAs (miR-10a, -99a, -184, -219 and -335) were selected for overexpression with the aim to induce differentiation in 2102Ep or NTera-2 cells. Of these, miR-184 induced a subtle mRNA downregulation of the SOX2 and NANOG pluripotency factors and an high upregulation of the marker of endodermal/trophectodermal differentiation. In NTera-2 cells, miR-335 upregulation caused a dramatic morphological change of the cells and the substantial downregulation of the OCT4 and NANOG mRNA transcripts. These results suggested that miR-184 and miR-335 could perform functions specific to 2102Ep and NTera-2 differentiation respectively. Computational target prediction produced OCT4 and SOX2 as potential targets for miR-335, both of which were downregulated at the protein level in miR-335-overexpressing NTera-2 cells. This showed for the first time that miR-335 could induce differentiation of an ECC line and that OCT4, SOX2 and NANOG were directly or indirectly regulated by this miRNA.

As a strong inducer of differentiation in a CSC model (NTera-2 cells), miR-335 could be an interesting therapeutic target for CSC-focussed treatment of cancer. Its tumour suppressive properties have been shown before in previous papers and, during this project, the interest in the

roles of miR-335 in cancer has rapidly grown. MiR-335 has also been implicated with mesenchymal and osteogenic differentiation.[370,372] So far, miR-335 has not yet been directly associated with CSC differentiation, although a very recent paper describes the contribution of miR-335 to the differentiation of glioma, an archetypal CSC disease, but does not specify or validate target genes relevant to this process.[397]

Interestingly, miR-335 upregulation in 2102Ep cells had no effect on cell differentiation or observable cell viability. In NTera-2 cells, this miRNA is upregulated by ATRA, suggesting it might be part of the retinoid-linked differentiation network.[274] Due to an aberrant differentiation program, 2102Ep cells are resistant to ATRA-induced differentiation, but also avoid terminal differentiation by other stimuli.[280,296,701] It is perhaps possible that miR-335 could be part of 2102Ep cells' defective differentiation mechanism and thus exert no influence on its state. On the other hand, previous published findings and the results in this thesis, especially those in Chapter 5, suggest that, 2102Ep and NTera-2 cells are quite dissimilar, despite sharing a germ cellular origin. Although both display common hESC markers, it is thought that 2102Ep cells more closely represent undifferentiated hESCs than NTera-2 cells, which represents a more committed cell type.[318,280,297] Previous studies have focussed on the role of miR-335 in adult stem cells, which resemble committed embryonic progenitor cells more closely than hESCs.[370,372] One study did show the upregulation of miR-335 in differentiating hESCs, but this was measured at seven days into the differentiation program, a stage at which cell commitment and specialisation is already prevailing.[702] No published results specifying an increase in miR-335 expression or that of its host gene, MEST, have been found in early differentiating hESCs or embryoid bodies. Considering these concepts, the hypothesis could be held that miR-335 is predominantly functional in more mature cells, rather than hESCs, inferring a possible cell-specific function in the NTera-2 cell line, while having no significant function in 2102Ep cells.

The results in Chapter 3 suggest that OCT4 and SOX2 are possible targets of miR-335. If these are indeed *in vivo* targets of this miRNA, compensatory, structural, regulatory or thermodynamic influences could perhaps explain the discrepancy between 2102Ep and NTera-2 response to miR-335 overexpression. Previous evidence that shows the context dependence of miRNA-target interaction does provide a possible answer for this discrepancy. Frequently, expressed miRNAs were found to regulate their expressed targets in certain cell lines, while, in other cell lines, this interaction was absent or compromised due its aiding or thwarting by mRNA binding elements and other miRNAs, but possibly also because of changes to the mRNA secondary structure.[197,211,703]

Of particular interest was the downregulation of SOX2 protein by miR-335, as the mRNA expression of this transcription factor remained high. This is a strong indicator of post-transcriptional regulation of SOX2, but it is unknown whether this is directly or indirectly mediated by miR-335. Interestingly, the downregulation of SOX2 protein was also measured during 3- and 6-day differentiation of NTera-2 cells with ATRA, while mRNA expression remains at high levels. The unchanged expression of SOX2 mRNA during the first six days of NTera-2 differentiation is characteristic for this cell line and is strongly supported by previous research, including results produced by this laboratory.[274,704,705] Still, one publication could be found which showed that both SOX2 mRNA and protein was downregulated seven days after ATRA-induced differentiation of NTera-2 cells, while another paper claims the downregulation of SOX2 mRNA after two days.[405,706] One study investigated the epigenetic transcriptional control of SOX2 in differentiated NTera-2 cells by DNA methylation and histone acetylation.[376] Through these mechanisms, SOX2 expression was found to be dynamically regulated in astrocytes but silenced in neurons. Also, a positive correlation was drawn between the expression of SOX2 and astrocyte proliferation, particularly in quiescent cells. This is interesting, because as quiescence is considered an important feature of CSC biology and is a suggested feature by which CSCs can avoid detection or treatment.[113] The ability of SOX2 to regulate cell proliferation of quiescent, more differentiated NTera-2 cells could link this gene to the lethality of these cells when transplanted into mouse brains.[707,708] Besides the control of SOX2 by DNA and histone modifications, the regulation of SOX2 by miRNA, as suggested in this thesis, might shed light on an additional epigenetic control mechanism through which SOX2 can exert its functions in CNS cells and cancers.

NCCIT cells, an ECC line that easily and terminally differentiates upon ATRA exposure, have previously been subjected to siRNA-induced SOX2 downregulation, which shows that this also causes them to differentiate.[378] In the nullipotent 2102Ep cell line, only OCT4 and NANOG have been downregulated previously, which resulted in the loss of ESC/ECC immunophenotype and the incomplete transformation into a more differentiated phenotype.[297,378] Also, NANOG downregulation in 2102Ep cells resulted in the loss of OCT4 and SOX2 expression, indicating the importance of this gene to the maintenance of stemness in this cell line.[378] SOX2 downregulation had not yet been tested in 2102Ep cells and with the goal to find novel ways to differentiate these cells, this gene was silenced in Chapter 4. As asserted after three days, SOX2 silencing resulted in the differentiation of the cells as measured by a reduction in pluripotency marker and AP expression and the upregulation of differentiation markers. In this laboratory, SOX2 silencing in 2102Ep cells was the first functional gene silencing experiment with siRNA in the ECC lines used for this study. Although SOX2 might not be an optimal therapeutic target, due to its prevalent

expression in many non-aberrant, mammalian adult cell types, the differentiation of 2102Ep cells by SOX2 knock-down proved to be a good test study to assess whether the siRNA-mediated gene knock down protocol designed could differentiate the cell lines as intended.

A similar experiment was performed in NTera-2 cells as the effects of SOX2 silencing in this cell line was unknown at the time. Like 2102Ep cells, NTera-2 cells appeared to differentiated upon SOX2 silencing. Interestingly, OCT4 expression remained high, three days after siSOX2 transfection, while NANOG significantly downregulated. After six days all pluripotency markers were downregulated, with SOX2 and NANOG almost silenced.

A previous study, published very recently by Gillis et al. (2011), also silenced SOX2 in NTera-2 cells upon which these appeared to differentiate as asserted by measuring the expression of pluripotency and differentiation markers.[405] Similarly to the results in this thesis, OCT4 mRNA expression analysis, performed by Gillis et al. with qRT-PCR, showed that OCT4 did not deregulate immediately. However, it is important to note here that OCT4 expression was measured two days after SOX2 silencing in NTera-2 cells compared to the three days of post-transfection incubation maintained in this thesis. Using qRT-PCR, Gillis et al. measured 31 genes of interest in SOX2-silenced NTera-2 cells. Although the expression of some genes remained unaltered in that study, only two genes (NANOG and HAND1) showed similar differential expression in the gene array data produced for this thesis. This is quite a large discrepancy, which could possibly be explained by the increased sensitivity of qRT-PCR over gene arrays, a difference in experimental conditions or design (e.g. higher siRNA concentration at lower cell density with a shorter incubation time used in a similar study by Gillis et al.), a phenotypic divergence between the NTera-2 strains used for either study or other external variables.[405]

2102Ep cells appear to respond to SOX2 downregulation in a similar way as hESCs. The more immediate response of OCT4 in this cell line to SOX2 downregulation, compared to NTera-2 cells, has also been found to be the case in hESCs, which already show a large downregulation of OCT4 upon SOX2 silencing after two days, which is reduced even more after four days.[459,709] It has to be noted that OCT4 was not downregulated in 2102Ep samples prepared for global transcriptome and miRNAome analysis. Although the reasons for this are unknown, but some hypothetical explanations are given in Section 5.5.1.1. It is clear that, to fully explain this incongruity, future repeats of the SOX2 knock down experiment are required, with perhaps different 2102Ep strains at different time points.

The core pluripotency marker-expressing 2102Ep and NTera-2 lineages both form lethal tumours after injection into mice.[37,707] However, previous studies have demonstrated that, when

xenografted into mice, they clearly differ in their capacity to proliferate and differentiate. Transplanted 2102Ep cells form a rapidly growing, highly undifferentiated tumour containing 100% ECCs, while NTera-2 cells generally form a well differentiated, tumour, with a histology representing many cell types and which can also be non-lethal, depending on the graft site.[37,296,313,707,708] This difference in cell behaviour could be pathologically related to a higher malignancy of 2102Ep cells compared to NTera-2 cells, as has been suggested before.[274,296] In short, the highly aggressive, largely undifferentiated teratocarcinomas formed by 2102Ep cells contain a high number of contributing CSCs, while the less aggressive, more differentiated NTera-2 teratocarcinomas contain relatively few CSCs. As an important mechanism for pluripotency and stem cell self-renewal, the SOX2 transcriptional network could potentially contribute to these characteristic phenotypes. In Chapter 5 the SOX2 transcriptional network of 2102Ep and NTera-2 cells were compared. This was the first time an extensive comparison of genes and miRNAs regulated by a member of a pluripotency network in these cell lines was performed.

These results showed that 2102Ep and NTera-2 cells respond quite differently to SOX2 downregulation. In 2102Ep cells, SOX2 knock down affected many genes and miRNAs that were common to undifferentiated and early differentiating hESCs. This phenotype matches the suggestion that 2102Ep cells are a good hESC model.[280] On the other hand, NTera-2 cells respond quite differently. Although a number of genes related to hESCs were deregulated, of which many were equally deregulated in 2102Ep cells, a large proportion of differentially regulated genes and miRNAs in NTera-2 cells were associated with a more committed progenitor phenotype. The downregulation of many neural markers suggest that SOX2 could have significant functions in maintaining a neural progenitor phenotype in NTera-2 cells. It is unclear whether undifferentiated NTera-2 populations consist of a mix of progenitor cells and cells of a more hESC phenotype, or whether a more homogenous phenotype exists with both hES and neural features. Recently, the NTera-2 cell line was reclassified to consist of two populations defined by the expression of Cripto-1.[600] It was found that cells with a high expression of Cripto-1 were more tumorigenic than those with low Cripto-1 expression. Additionally, NTera-2 cells with high Cripto-1 expression also had higher OCT4, SOX2 and NANOG expression, while in the low Cripto-1 expressing cells, neural markers, among other dermal markers, were enriched. It is possible that SOX2 downregulation conferred specific alterations in these different populations, resulting in a mixed expression profile. The high enrichment of committed progenitor markers in the NTera-2 differential gene and miRNA data could suggest the dominance of a less tumorigenic phenotype in the population over the more tumorigenic, ESC-like phenotype. Consequently, the very high tumorigenicity of 2102Ep cells could be related to a phenotype that, inferred from the differential gene and miRNA data, could be even closer to that of hESCs.

Cell potency, as a phenotypic trait of all stem cells, has been considered as an important factor in CSCs and cancer. Some CSCs have characteristically high expression of pluripotency factors OCT4, SOX2 and NANOG, particularly germ cell cancers, such as embryonal carcinoma, which, in contrast to normal adult germ cell lines, can have pluripotent traits.[282] Although pluripotency is not a general characteristic of most CSCs, pluripotency factors have been found to be expressed in several stem cell-like cancer cells. Prior to this study, SOX2 has been associated, sometimes alongside OCT4, with germ cell tumours and cancers of the gastrointestinal tract, prostate and the CNS.[57,282,508,710–714] Contradictorily, SOX2 has also been shown to have tumour suppressive functions in gastric cancers, with reported downregulation in various cases of this disease.[670,715] Initial research has been focussed on SOX2 as a component of ESC and ECC pluripotency. Recent research, further detailed in the introductions and discussions of the previous chapters, has now shown SOX2 to have supplementary, but crucial roles in the CNS, adult stem cell biology and cancer.

The Wnt signalling pathway is one of the most analysed pathways in cancer research and has highly documented functions in CSCs. This has been extensively described in the introduction of this thesis. The overrepresentation of the Wnt pathway among the differential gene expression data from 2102Ep and NTera-2 cells, and the presence of potential SOX2 binding sites in the promoters of most of the deregulated members of this pathway, appears to suggest that SOX2 is an important regulator of Wnt signalling in ECCs. In both 2102Ep and NTera-2 cells, a larger number of Wnt pathway-related genes are upregulated than downregulated. The transcriptional activation of the Wnt signalling elements has previously been observed *in vivo* in histologically differentiated non-seminomas compared to embryonal carcinoma.[442] With the findings from Chapter 5, this observation provides evidence to suggest that SOX2 could be a central regulator of the Wnt pathway in ECCs and perhaps other SOX2-expressing CSCs. When focused more closely on the individual components of Wnt signalling in 2102Ep and NTera-2 cells, it was striking to see that DKK1, an important Wnt pathway antagonist and potent tumour suppressor, was among the highest upregulated genes in both cell lines, implying the function of SOX2 as a transcriptional suppressor of this gene in ECCs. DKK1 is typically a gene that is highly upregulated upon ESC or ECC differentiation by ATRA and is required for neural differentiation, but its upregulation has not been shown in 2102Ep or NTera-2 cells after exposing them to ATRA.[536,579,705,716,717] Conversely, a very recent study showed that, in hMSCs, SOX2 is a direct transcriptional activator of DKK1, which in turn, upon its overexpression, results in a strongly altered hMSC differentiation pattern, affecting the cell fate of these multipotent stem cells.[588] The aforementioned study is the only one found that reveals an important regulatory link between SOX2 and DKK1 in adult stem cells, but given the increasingly documented roles of both SOX2 and DKK1 in other adult stem cell types,

notably neural stem cells, it is quite possible that the regulatory relationship of SOX2 and DKK1 is more widespread. Section 5.5.1.2 details the role of DKK1 as an important activator of the CNS cell fate in stem cells. If SOX2 retains its role as a transcriptional regulator of DKK1 throughout the ESC/ECC's differentiation program, this could mean that somewhere during the differentiation of these cells to CNS cells, SOX2 switches from a transcriptional repressor of DKK1 to an activator. It is unclear whether the reduction of SOX2 protein in NTera-2 cells, as shown in Section 3.4.6.6, is part of this switching process. This may be worth investigating and because of the potent anti-carcinogenic trait of DKK1 and the oncogenic role of SOX2, the differential relationship of these two genes could be an interesting target of further research in cancer and CSC biology.

Previous studies demonstrated that NTera-2 cells become particularly tumourigenic, creating very lethal undifferentiated tumours, when transplanted in certain anatomical parts of the murine brain.[707,708] When transplanted to other locations in the brain, NTera-2 cells were less lethal and differentiated into neuronal tissue.[707,708] While this is interesting supporting evidence for the microenviromental importance to CSC oncogenicity, given the central role of SOX2 in embryonic and adult neural stem cell maintenance, the hypothesis could be held that the impact of the microenvironment on the tumourigenicity of NTera-2 cells might be channelled through the SOX2 network. However, this cannot be confirmed without repeating the xenotransplantation experiments with a focus on this network.

A large number of miRNAs with previously documented functions in CSCs were deregulated in 2102Ep and NTera-2 cells by SOX2 downregulation. Some of the cancers of origin have been shown to have elevated, and in some cases downregulated levels of SOX2 expression when compared to normal tissue. The substantial number of CSC-related miRNAs deregulated by SOX2, as seen in Section 5.5.2.1, gives rise to the possibility that this transcription factor orchestrates carcinogenic miRNA expression. The downregulation of miR-372 in NTera-2 cells stood out, as this miRNA has previously been shown by Voorhoeve et al. (2006) to convey oncogenicity in TGCTs by upregulating the embryonal carcinoma component of these cancers.[597] MiR-372 is a member of the oncogenic miR-371-373 cluster, of which its members have been detected in undifferentiated 2102Ep and NTera-2 cells by this study and by a previous study from this laboratory.[274] Upon SOX2 silencing, its other members (miR-371-3p and miR-373) were also significantly downregulated, albeit slightly below the 2-fold cut off in NTera-2 cells and, including miR-372, far lower than 2-fold in 2102Ep cells. A very recent study showed that the Wnt signalling pathway transcriptionally activates the miR-371-373 cluster through β-catenin/LEF1 in different cancer cell lines, resulting in the suppression of DKK1 by miR-372 and miR-373.[580] This mechanism is probably occurring in 2102Ep and NTera-2 cells, and shows the integral position of SOX2 in this network. The findings

described in this thesis succinctly tie together the pluripotent stem-like nature of germ cell CSCs and the biology of other somatic cancer cell lines.

In Section 5.3.3.4, a computational and statistical method was used to cross-reference the differential miRNA and gene expression data produced for Chapter 5. For 2102Ep cells, this yielded 15 genes that were oppositely regulated compared to the median expression of the group of differentially expressed miRNAs that were computationally predicted to target them. Although the computational method used for the prediction of these miRNA:target relationships was based on a previous study, it is uncertain that this is valid without conformational studies. From this, a few potential targets stand out. In particular, LEF1 and DKK1, up- and downstream components of the Wnt signalling pathway respectively. These were predicted to be predominantly regulated by miRNAs of the C19MC polycistron. Twenty-eight members of this miRNA cluster were downregulated in 2102Ep cells after SOX2 knock down, strongly suggesting a functional significance during 2102Ep cell differentiation. It has also frequently been found to be enriched in undifferentiated hESCs, but an extensive characterisation of this polycistron has not yet been performed.[175,687,688,690] It strong expression in undifferentiated 2102Ep cells and the nearly-complete absence of this in NTera-2 cells signify the difference between these two cell lines. The C19MC miRNAs also share their seed sequence with various other miRNAs elsewhere in the genome. The important roles of some of these miRNAs in hESC and cancer biology might suggest significant cross-functionality exists between C19MC and the miRNA networks that govern these processes. Indeed, although research is still scarce, the oncogenic nature of C19MC members have previously been shown.[691–693] Together, the C19MC miRNAs downregulated in 2102Ep cells have been computationally predicted to target almost a quarter of all upregulated genes, many of which are important in stem cell regulation and cancer. For example, many of these are upregulated members of the Wnt signalling pathway. This might signify the potential importance of the C19MC cluster in 2102Ep cells and the exclusivity of this cluster over NTera-2 cells could be indicative of a tumourigenic difference between these cell lines.

Finally, although not detected in the promoter region of the miR-371-373 or C19MC clusters, the promoters of many deregulated genes and miRNAs contain SOX2 binding sites validated in other cell types. Notably, some deregulated miRNAs in 2102Ep and NTera-2 cells were previously discovered to be post-transcriptional regulators of SOX2. SOX2 was a previously validated target of miR-126, miR-145 (both downregulated in NTera-2 cells) and miR-9* (upregulated in 2102Ep cells), of which the latter two have SOX2 binding sites in their promoters, as shown in mESCs.[270,466,523,670] Although, the existence of a miR-145-SOX2 regulatory loop has been proposed in GBM, an autoregulatory relationship between SOX2 and miR-126 and miR-9* has yet to be

established.[466] As described in Section 5.5.2.2, miR-9* and miR-145 are both upregulated in early differentiated ESCs and both have tumour suppressive functions in neoplastic cells. Profiling the SOX2-miRNA autoregulatory network could expand the knowledge of miRNA regulation in ESCs, but could also provide potentially valuable targets for cancer therapy as these miRNAs could possibly be targeted as endogenous repressors of the SOX2 oncogene in some malignancies.

6.2 — Prospects for Future Work

The research performed in this study and discussed in this dissertation has yielded some interesting results that could be carried forward to future research projects. Some of the results require little further experimentation or validation to pursue the hypotheses brought forward in the discussions of this thesis. The functional miRNA experiments in Chapter 3 produced two miRNAs of interest for future research, miR-335 and miR-184. MiR-335 has shown a promising capability of differentiating NTera-2 cells resulting in the downregulation of OCT4 and NANOG transcripts and OCT4 and SOX2 protein (NANOG protein expression was not screened). One of the most enticing experiments to conduct, would be a validation study with the aim of confirming OCT4 and SOX2 as targets of miR-335. For these, SOX2 is a particularly interesting candidate as it is silenced at the protein level, but not transcriptionally, strongly suggesting post-transcriptional regulation of this gene is taking place. MiRNA target validation is commonly performed with the widely used luciferase reporter assays (LRAs). With these, miRNA:target relationships can be verified by cloning the 3' UTR of the target into a luciferase vector. Both the 3' UTR-luciferase vector and the candidate miRNA are transfected into cells after which a reduced fluorescence from expressed luciferase is detected if the miRNA represses the vector's expression by binding on the 3' UTR target site. When combined with the correct negative controls, including a 3' UTR-luciferase vector with a scrambled putative miRNA target site, LRAs are very valuable and reliable tools to establish miRNA:target relationships.

Additionally, in the context of CSC research, it would be recommendable to further assess the *in vitro* effects of miR-335 overexpression in ECCs and perhaps other CSCs on self-renewal and metastasis at predetermined time points after miRNA upregulation. For this, cell proliferation experiments (e.g. MTT assay) or invasion experiments (e.g. Matrigel invasion assay) would be required. These are relatively simple experiments to perform, but could shed light on more specific tumour suppressive functions of miR-335 in CSCs. Further down the line, mouse xenografting experiments with miR-335 overexpressing ECCs and other CSCs could assess whether miR-335 has similar CSC suppressive properties *in vivo.* A novel method for performing this experiment could be to engineer a ECC or CSC line with a miR-335 overexpression or inducible MEST/miR-335 gene (e.g. by doxycyclin), which could subsequently be xenografted in a model organism, such as immunocompromised mice. MiR-335 overexpression in NTera-2 cells could help assess its potential to inhibit tumour initiation by this cell line. Although overexpressing miRNAs in cells prior to their xenotransplanation is common, the method of doxycyclin-induced *in vivo* gene activation has been tested before and would, once a protocol is established, additionally

allow for an easy way to regulate the ectopic *in vivo* expression of miR-335 study its effects at chosen time-points during cancer development, such as during the metastasis phase.[718]

Although not a inducer of differentiation, miR-184 did have some influence on pluripotency/differentiation marker expression in 2102Ep cells and therefore could still be a player in the differentiation program of this cell line. TCF7L1 was suggested as a miR-184 target in Section 3.5.3. As detailed in this Section, TCF7L1, also frequently referred to as Tcf3, is a transcription factor that is downstream from the Wnt and TGFβ signalling pathways and regularly teams up with LEF to regulate gene expression during embryonic development and stem cell differentiation. In the context of the results from Chapter 5, which imply the importance of Wnt and TGFβ signalling in 2102Ep cells, TCF7L1 could be an interesting target of miR-184 to pursue for testing and validation.

The foundational experiments in Chapter 4 and 5 showed that SOX2 knock down conferred an early differentiated phenotype in 2102Ep and NTera-2 cells, which upon further investigation, produced differential gene and miRNA expression results that could be used for more focussed studies. Before the research for this thesis was performed, little was known about SOX2 function in 2102Ep cells and no functional analysis was performed with this gene in this cell line. 2102Ep cells, as a nullipotent cell line, has, so far, been shown to be quite resistant to spontaneous or morphogen-induced differentiation. As shown previously, upon the knock-down of OCT4, 2102Ep cells do differentiate to the trophectoderm.[297] However, this study was only performed for five days after initial OCT4 knock-down. Unpublished data from the cell line's original founder, Peter Andrews, showed that OCT4 knock-down could not terminally differentiate 2102Ep cells. The results from Chapter 5 of this thesis show that 2102Ep cells do differentiate after the knock-down of SOX2. An interesting follow-up study could investigate the long-term fate of 2102Ep cells after the downregulation of this gene. This could reveal if 2102Ep cells, or a subpopulation of this cell line could regain the capacity to differentiate or if a senescent relatively unspecified cell type prevails. For this, it would be wise to either pursue a method to uniformly deactivate SOX2 in 2102Ep cells (e.g. by doxycylin-inducible deactivation of this gene) or to remove any SOX2-expressing subpopulations from the culture by e.g. fluorescent-activated cell sorting flow cytometry. Either of these methods would have to be used, otherwise SOX2-expressing 2102Ep cells would be allowed to rapidly proliferate and dominate the culture.

Because of the similarities, but also many differences in differential gene and miRNA expression between 2102Ep and NTera-2 cells a ChIP analysis with the promoter regions of selected genes and miRNAs could confirm direct transcriptional regulation of these by SOX2. Due to their significance to cancer and embryonic development, members of the Wnt and TGFβ signalling

network with putative SOX2 binding sites could be validated as transcriptional targets of SOX2. For example, the validation of DKK1 expression by SOX2 in 2102Ep and NTera-2 cells could implicate the DKK1-SOX2 autoregulatory mechanism in CSCs, while a further exploration of SOX2 binding to the DKK1 promoter during NTera-2 cell differentiation could establish temporal alterations between the repressive or inductive function of SOX2 to DKK1 expression. Alternatively, mapping the entire genome for SOX2 binding sites by a ChIP-sequencing study could reveal novel transcriptional relationships between SOX2 and target genes, some of which might be exclusive to 2102Ep or NTera-2 cells. This would complete the grand picture of the SOX2 transcriptional network in two common ECC lines.

Similarly, with the findings from Chapter 5, additional transcriptional and posttranscriptional SOX2:miRNA relationships could be established. The proposed autoregulatory SOX2:miRNA loops would be interesting candidates because of their direct role in SOX2, and by effect, pluripotency marker homeostasis. Another interesting focus for future research would be the characterisation of the C19MC miRNAs in the context of ECC and CSC biology, as their association to (cancer) stemness not has been well described, despite their reported oncogenic potential and their enrichment in ESCs.[175,687–693] Because many of the C19MC miRNAs have been predicted in Section 5.4.7.5 to target several interesting genes, such as Wnt pathway members DKK1 and LEF1, a validation of these could ascribe novel developmental and oncogenic functions to this C19MC and provide a level validation of the miRNA target prediction method used in this section.

REFERENCES

References

1. Pierce, G. B. Neoplasms, differentiations and mutations. *Am. J. Pathol.* **77**, 103–118 (1974).
2. Huntly, B. J. P. & Gilliland, D. G. Leukaemia stem cells and the evolution of cancer-stem-cell research. *Nat Rev Cancer* **5**, 311–321 (2005).
3. Virchow, R. Virchows Archiv für pathologische Anatomie und Physiologie und für klinische Medizin. *Virchows Archiv für pathologische Anatomie und Physiologie und für klinische Medizin* **23**, 23 (1855).
4. Paget, S. THE DISTRIBUTION OF SECONDARY GROWTHS IN CANCER OF THE BREAST. *The Lancet* **133**, 571–573 (1889).
5. Récamier, J. C. . *Recherches sur le Traitement du Cancer*. (Paris, 1829).
6. Cohnheim, J. Ueber Entzündung und Eiterung. *Archiv für Pathologische Anatomie und Physiologie und für Klinische Medicin* **40**, 1–79 (1867).
7. Durante, F. Nesso fisio-pathologico tra la struttura dei nei materni e la genesi di alcuni tumori maligni. *Arch Memori ed Osservazioni di Chirugia Practica* **11**, 217–226 (1874).
8. Bainbridge, W. S. *The Cancer problem*. (Macmillan: 1914).
9. Sell, S. On the Stem Cell Origin of Cancer. *The American Journal of Pathology* **176**, 2584–2594 (2010).
10. PIERCE, B., VERNEY, E. L. & DIXON, F. J. The biology of testicular cancer. I. Behavior after transplantation. *Cancer Res.* **17**, 134–138 (1957).
11. Southam, C. M. & Brunschwig, A. Quantitative studies of autotransplantation of human cancer. Preliminary report. *Cancer* **14**, 971–978 (1961).
12. BRUCE, W. R. & VAN DER GAAG, H. A QUANTITATIVE ASSAY FOR THE NUMBER OF MURINE LYMPHOMA CELLS CAPABLE OF PROLIFERATION IN VIVO. *Nature* **199**, 79–80 (1963).
13. Wodinsky, I., Swiniarski, J. & Kensler, C. Spleen colony studies of leukemia L1210. I. Growth kinetics of lymphocytic L1210 cells in vivo as determined by spleen colony assay. *Cancer Chemother. Rep* **51**, 415–421 (1967).
14. Bergsagel, D. E. & Valeriote, F. A. Growth characteristics of a mouse plasma cell tumor. *Cancer Res.* **28**, 2187–2196 (1968).
15. Becker, A. J., McCulloch, E. A. & Till, J. E. Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells. *Nature* **197**, 452–454 (1963).
16. Siminovitch, L., McCulloch, E. A. & Till, J. E. The distribution of colony-forming cells among spleen colonies. *J. Cell. Physiol.* **62**, 327–336 (1963).
17. Park, C. H., Bergsagel, D. E. & McCulloch, E. A. Mouse myeloma tumor stem cells: a primary cell culture assay. *J. Natl. Cancer Inst.* **46**, 411–422 (1971).
18. Stevens, L. C. Experimental production of testicular teratomas in mice. *Proceedings of the National Academy of Sciences of the United States of America* **52**, 654 (1964).
19. Stevens, L. C. Origin of testicular teratomas from primordial germ cells in mice. *J. Natl. Cancer Inst.* **38**, 549–552 (1967).
20. PIERCE, G. B., Jr, DIXON, F. J., Jr & VERNEY, E. L. Teratocarcinogenic and tissue-forming potentials of the cell types comprising neoplastic embryoid bodies. *Lab. Invest.* **9**, 583–602 (1960).
21. KLEINSMITH, L. J. & PIERCE, G. B., Jr MULTIPOTENTIALITY OF SINGLE EMBRYONAL CARCINOMA CELLS. *Cancer Res.* **24**, 1544–1551 (1964).
22. Finch, B. W. & Ephrussi, B. RETENTION OF MULTIPLE DEVELOPMENTAL POTENTIALITIES BY CELLS OF A MOUSE TESTICULAR TERATOCARCINOMA DURING PROLONGED CULTURE in vitro AND THEIR EXTINCTION UPON HYBRIDIZATION WITH CELLS OF PERMANENT LINES. *Proc. Natl. Acad. Sci. U.S.A.* **57**, 615–621 (1967).
23. Kahan, B. W. & Ephrussi, B. Developmental potentialities of clonal in vitro cultures of mouse testicular teratoma. *J. Natl. Cancer Inst.* **44**, 1015–1036 (1970).

24. Fox, N., DeSouza, L., Simon, D. & Damjanov, I. Male murine embryonal carcinoma cell line selectively metastatic to the ovaries and adrenals. *Virchows Arch., B, Cell Pathol.* **43**, 241–251 (1983).
25. McBurney, M. W. & Rogers, B. J. Isolation of male embryonal carcinoma cells and their chromosome replication patterns. *Dev. Biol.* **89**, 503–508 (1982).
26. McBurney, M. W., Jones-Villeneuve, E. M., Edwards, M. K. & Anderson, P. J. Control of muscle and neuronal differentiation in a cultured embryonal carcinoma cell line. *Nature* **299**, 165–167 (1982).
27. Martin, G. R. & Evans, M. J. The morphology and growth of a pluripotent teratocarcinoma cell line and its derivatives in tissue culture. *Cell* **2**, 163–172 (1974).
28. Jakob, H., Boon, T., Gaillard, J., Nicolas, J. & Jacob, F. [Teratocarcinoma of the mouse: isolation, culture and properties of pluripotential cells]. *Ann. Microbiol. (Paris)* **124**, 269–282 (1973).
29. Martin, G. R. & Evans, M. J. Differentiation of clonal lines of teratocarcinoma cells: formation of embryoid bodies in vitro. *Proc. Natl. Acad. Sci. U.S.A.* **72**, 1441–1445 (1975).
30. Evans, M. J. The isolation and properties of a clonal tissue culture strain of pluripotent mouse teratoma cells. *J Embryol Exp Morphol* **28**, 163–176 (1972).
31. Nicolas, J. F., Dubois, P., Jakob, H., Gaillard, J. & Jacob, F. [Mouse teratocarcinoma: differentiation in cultures of a multipotential primitive cell line (author's transl)]. *Ann. Microbiol. (Paris)* **126**, 3–22 (1975).
32. Berstine, E. G., Hooper, M. L., Grandchamp, S. & Ephrussi, B. Alkaline phosphatase activity in mouse teratoma. *Proc. Natl. Acad. Sci. U.S.A.* **70**, 3899–3903 (1973).
33. Bernstine, E. G., Hooper, M. L., Grandchamp, S. & Ephrussi, B. Alkaline Phosphatase Activity in Mouse Teratoma. *Proc Natl Acad Sci U S A* **70**, 3899–3903 (1973).
34. Fogh, J. New human tumor cell lines. *Human tumor cells in vitro* 115–159 (1975).
35. Hogan, B., Fellous, M., Avner, P. & Jacob, F. Isolation of a human teratoma cell line which expresses F9 antigen. *Nature* **270**, 515–518 (1977).
36. Andrews, P. W. From teratocarcinomas to embryonic stem cells. *Philosophical Transactions of the Royal Society B: Biological Sciences* **357**, 405–417 (2002).
37. Andrews, P. W. *et al.* Pluripotent embryonal carcinoma clones derived from the human teratocarcinoma cell line Tera-2. Differentiation in vivo and in vitro. *Lab. Invest* **50**, 147–162 (1984).
38. Pera, M. F., Cooper, S., Mills, J. & Parrington, J. M. Isolation and characterization of a multipotent clone of human embryonal carcinoma cells. *Differentiation* **42**, 10–23 (1989).
39. Teshima, S. *et al.* Four new human germ cell tumor cell lines. *Lab. Invest.* **59**, 328–336 (1988).
40. Damjanov, I., Horvat, B. & Gibas, Z. Retinoic acid-induced differentiation of the developmentally pluripotent human germ cell tumor-derived cell line, NCCIT. *Lab. Invest.* **68**, 220–232 (1993).
41. Andrews, P. W., Goodfellow, P. N., Shevinsky, L. H., Bronson, D. L. & Knowles, B. B. Cell-surface antigens of a clonal human embryonal carcinoma cell line: morphological and antigenic differentiation in culture. *Int. J. Cancer* **29**, 523–531 (1982).
42. Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154–156 (1981).
43. Martin, G. R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl. Acad. Sci. U.S.A.* **78**, 7634–7638 (1981).
44. Thomson, J. A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145–1147 (1998).
45. Pierce, G. B. & Johnson, L. D. Differentiation and Cancer. *In Vitro* **7**, 140–145 (1971).
46. Potter, V. R. Phenotypic diversity in experimental hepatomas: the concept of partially blocked ontogeny. The 10th Walter Hubert Lecture. *Br J Cancer* **38**, 1–23 (1978).
47. MAKINO, S. Further evidence favoring the concept of the stem cell in ascites tumors of rats. *Ann. N. Y. Acad. Sci.* **63**, 818–830 (1956).

48. Pierce, G. B. & Wallace, C. Differentiation of malignant to benign cells. *Cancer Res.* **31**, 127–134 (1971).
49. Pierce, G. B., Shikes, R. & Fink, L. M. *Cancer: a problem of developmental biology*. (Prentice-Hall: 1978).
50. Raghavan, D. & Society, A. C. *Germ Cell Tumors*. (2003).
51. Hamburger, A. W. & Salmon, S. E. Primary bioassay of human tumor stem cells. *Science* **197**, 461–463 (1977).
52. Martinez-Climent, J. A., Fontan, L., Gascoyne, R. D., Siebert, R. & Prosper, F. Lymphoma stem cells: enough evidence to support their existence? *Haematologica* **95**, 293–302 (2010).
53. Lapidot, T. *et al.* A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature* **367**, 645–648 (1994).
54. Bonnet, D. & Dick, J. E. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat. Med.* **3**, 730–737 (1997).
55. Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J. & Clarke, M. F. Prospective identification of tumorigenic breast cancer cells. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3983–3988 (2003).
56. Singh, S. K. *et al.* Identification of a cancer stem cell in human brain tumors. *Cancer Res.* **63**, 5821–5828 (2003).
57. Hemmati, H. D. *et al.* Cancerous stem cells can arise from pediatric brain tumors. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 15178–15183 (2003).
58. Ignatova, T. N. *et al.* Human cortical glial tumors contain neural stem-like cells expressing astroglial and neuronal markers in vitro. *Glia* **39**, 193–206 (2002).
59. O'Brien, C. A., Pollett, A., Gallinger, S. & Dick, J. E. A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. *Nature* **445**, 106–110 (2007).
60. Zhang, S. *et al.* Identification and characterization of ovarian cancer-initiating cells from primary human tumors. *Cancer Res.* **68**, 4311–4320 (2008).
61. Li, C. *et al.* Identification of pancreatic cancer stem cells. *Cancer Res.* **67**, 1030–1037 (2007).
62. Prince, M. E. *et al.* Identification of a subpopulation of cells with cancer stem cell properties in head and neck squamous cell carcinoma. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 973–978 (2007).
63. Fang, D. *et al.* A tumorigenic subpopulation with stem cell properties in melanomas. *Cancer Res.* **65**, 9328–9337 (2005).
64. Lang, S., Frame, F. & Collins, A. Prostate cancer stem cells. *J Pathol* **217**, 299–306 (2009).
65. Lee, E.-K., Cho, H. & Kim, C.-W. Proteomic analysis of cancer stem cells in human prostate cancer cells. *Biochem. Biophys. Res. Commun.* **412**, 279–285 (2011).
66. Manton, K. G., Akushevich, I. & Kravchenko, J. *Cancer Mortality and Morbidity Patterns in the U.S. Population: An Interdisciplinary Approach*. (Springer: 2009).
67. Marcum, J. A. Metaphysical presuppositions and scientific practices: Reductionism and organicism in cancer research. *International Studies in the Philosophy of Science* **19**, 31–45 (2005).
68. Trosko, J. E. From Adult Stem Cells to Cancer Stem Cells. *Annals of the New York Academy of Sciences* **1089**, 36–58 (2006).
69. Shipitsin, M. & Polyak, K. The cancer stem cell hypothesis: in search of definitions, markers, and relevance. *Lab Invest* **88**, 459–463 (2008).
70. Farrar, W. L. & Hurt, E. Characterization Of Cancer Stem Cells. *Cancer stem cells* 1–49 (2009).
71. Scatena, R., Bottoni, P., Pontoglio, A. & Giardina, B. Cancer stem cells: the development of new cancer therapeutics. *Expert Opinion on Biological Therapy* **11**, 875–892 (2011).
72. Farrar, W. L. *Cancer stem cells*. (Cambridge University Press: 2009).
73. Li, L. & Neaves, W. B. Normal Stem Cells and Cancer Stem Cells: The Niche Matters. *Cancer Research* **66**, 4553 –4557 (2006).
74. Baumann, M., Krause, M. & Hill, R. Exploring the role of cancer stem cells in radioresistance. *Nat Rev Cancer* **8**, 545–54 (2008).
75. Nekulova, M., Holcakova, J., Coates, P. & Vojtesek, B. The role of p63 in cancer, stem cells and cancer stem cells. *Cell. Mol. Biol. Lett.* **16**, 296–327 (2011).

76. Shay, J. W. & Wright, W. E. Telomeres and telomerase in normal and cancer stem cells. *FEBS Letters* **584**, 3819–3825 (2010).
77. Gilbert, D., Rapley, E. & Shipley, J. Testicular germ cell tumours: predisposition genes and the male germ cell niche. *Nat. Rev. Cancer* **11**, 278–288 (2011).
78. Kashyap, V. *et al.* Regulation of stem cell pluripotency and differentiation involves a mutual regulatory circuit of the NANOG, OCT4, and SOX2 pluripotency transcription factors with polycomb repressive complexes and stem cell microRNAs. *Stem Cells Dev.* **18**, 1093–1108 (2009).
79. Boyer, L. A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947–956 (2005).
80. Ng, H.-H. & Surani, M. A. The transcriptional and signalling networks of pluripotency. *Nat. Cell Biol* **13**, 490–496 (2011).
81. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
82. Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917–1920 (2007).
83. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
84. Baker, M. Adult cells reprogrammed to pluripotency, without tumors. *Nature Reports Stem Cells* (2007).doi:10.1038/stemcells.2007.124
85. Knoepfler, P. S. Deconstructing stem cell tumorigenicity: a roadmap to safe regenerative medicine. *Stem Cells* **27**, 1050–1056 (2009).
86. Ben-Porath, I. *et al.* An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nat Genet* **40**, 499–507 (2008).
87. Wong, D. J. *et al.* Module map of stem cell genes guides creation of epithelial cancer stem cells. *Cell Stem Cell* **2**, 333–344 (2008).
88. Chaffer, C. L. & Weinberg, R. A. A perspective on cancer cell metastasis. *Science* **331**, 1559–1564 (2011).
89. Poste, G. & Fidler, I. J. The pathogenesis of cancer metastasis. *Nature* **283**, 139–146 (1980).
90. Ramaswamy, S., Ross, K. N., Lander, E. S. & Golub, T. R. A molecular signature of metastasis in primary solid tumors. *Nat. Genet.* **33**, 49–54 (2003).
91. Webb, T. Microarray Studies Challenge Theories of Metastasis. *Journal of the National Cancer Institute* **95**, 350 –351 (2003).
92. Ghadially, R. The role of stem and circulating cells in cancer metastasis. *Journal of Surgical Oncology* **103**, 555–557 (2011).
93. Armstrong, A. J. *et al.* Circulating Tumor Cells from Patients with Advanced Prostate and Breast Cancer Display Both Epithelial and Mesenchymal Markers. *Molecular Cancer Research* **9**, 997 –1007 (2011).
94. Klonisch, T. *et al.* Cancer stem cell markers in common cancers - therapeutic implications. *Trends Mol Med* **14**, 450–460 (2008).
95. Ocana, O. H. & Nieto, M. A. Epithelial plasticity, stemness and pluripotency. *Cell Res* **20**, 1086–1088 (2010).
96. Fidler, I. J. & Kripke, M. L. Metastasis results from preexisting variant cells within a malignant tumor. *Science* **197**, 893–895 (1977).
97. Rosen, J. M. & Jordan, C. T. The Increasing Complexity of the Cancer Stem Cell Paradigm. *Science* **324**, 1670 –1673 (2009).
98. Brabletz, T., Jung, A., Spaderna, S., Hlubek, F. & Kirchner, T. Migrating cancer stem cells - an integrated concept of malignant tumour progression. *Nat Rev Cancer* **5**, 744–749 (2005).
99. Kong, D., Li, Y., Wang, Z. & Sarkar, F. H. Cancer Stem Cells and Epithelial-to-Mesenchymal Transition (EMT)-Phenotypic Cells: Are They Cousins or Twins? *Cancers (Basel)* **3**, 716–729 (2011).
100. Hermann, P. C. *et al.* Distinct Populations of Cancer Stem Cells Determine Tumor Growth and Metastatic Activity in Human Pancreatic Cancer. *Cell Stem Cell* **1**, 313–323 (2007).

101. Pang, R. *et al.* A Subpopulation of CD26+ Cancer Stem Cells with Metastatic Capacity in Human Colorectal Cancer. *Cell Stem Cell* **6**, 603–615 (2010).
102. Morel, A.-P. *et al.* Generation of breast cancer stem cells through epithelial-mesenchymal transition. *PLoS ONE* **3**, e2888 (2008).
103. Mani, S. A. *et al.* The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* **133**, 704–715 (2008).
104. Santisteban, M. *et al.* Immune-induced epithelial to mesenchymal transition in vivo generates breast cancer stem cells. *Cancer Res.* **69**, 2887–2895 (2009).
105. Gupta, P. B. *et al.* Identification of selective inhibitors of cancer stem cells by high-throughput screening. *Cell* **138**, 645–659 (2009).
106. Liu, H.-G., Chen, C., Yang, H., Pan, Y.-F. & Zhang, X.-H. Cancer stem cell subsets and their relationships. *J Transl Med* **9**, 50 (2011).
107. Kalluri, R. & Weinberg, R. A. The basics of epithelial-mesenchymal transition. *Journal of Clinical Investigation* **119**, 1420–1428 (2009).
108. Paterlini-Brechot, P. & Benali, N. L. Circulating tumor cells (CTC) detection: clinical impact and future directions. *Cancer Lett.* **253**, 180–204 (2007).
109. Mocellin, S., Keilholz, U., Rossi, C. R. & Nitti, D. Circulating tumor cells: the 'leukemic phase' of solid cancers. *Trends in Molecular Medicine* **12**, 130–139 (2006).
110. Kalluri, R. EMT: When epithelial cells decide to become mesenchymal-like cells. *Journal of Clinical Investigation* **119**, 1417–1419 (2009).
111. Thiery, J. P. Epithelial-mesenchymal transitions in tumour progression. *Nat Rev Cancer* **2**, 442–454 (2002).
112. Clarke, M. F. *et al.* Cancer stem cells--perspectives on current status and future directions: AACR Workshop on cancer stem cells. *Cancer Res* **66**, 9339–44 (2006).
113. Wang, J. C. Y. Evaluating Therapeutic Efficacy against Cancer Stem Cells: New Challenges Posed by a New Paradigm. *Cell Stem Cell* **1**, 497–501 (2007).
114. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
115. Dittmar, T. *et al.* Recurrence cancer stem cells – Made by cell fusion? *Medical Hypotheses* **73**, 542–547 (2009).
116. Deonarain, M. P., Kousparou, C. A. & Epenetos, A. A. Antibodies targeting cancer stem cells: a new paradigm in immunotherapy? *MAbs* **1**, 12–25 (2009).
117. Dean, M., Fojo, T. & Bates, S. Tumour stem cells and drug resistance. *Nat Rev Cancer* **5**, 275–284 (2005).
118. Rich, J. N. Cancer Stem Cells in Radiation Resistance. *Cancer Research* **67**, 8980 –8984 (2007).
119. Heddleston, J. M. *et al.* Hypoxia inducible factors in cancer stem cells. *Br J Cancer* **102**, 789–795 (2010).
120. Lathia, J. D., Heddleston, J. M., Venere, M. & Rich, J. N. Deadly Teamwork: Neural Cancer Stem Cells and the Tumor Microenvironment. *Cell Stem Cell* **8**, 482–485 (2011).
121. Kawasaki, B. T., Mistree, T., Hurt, E. M., Kalathur, M. & Farrar, W. L. Co-expression of the toleragenic glycoprotein, CD200, with markers for cancer stem cells. *Biochemical and Biophysical Research Communications* **364**, 778–782 (2007).
122. Kawasaki, B. T. & Farrar, W. L. Cancer stem cells, CD200 and immunoevasion. *Trends in Immunology* **29**, 464–468 (2008).
123. Jaiswal, S. *et al.* CD47 is upregulated on circulating hematopoietic stem cells and leukemia cells to avoid phagocytosis. *Cell* **138**, 271–285 (2009).
124. Yaguchi, T. *et al.* The mechanisms of cancer immunoescape and development of overcoming strategies. *International Journal of Hematology* **93**, 294–300 (2011).
125. Wu, A. *et al.* Glioma cancer stem cells induce immunosuppressive macrophages/microglia. *Neuro-oncology* **12**, 1113–1125 (2010).
126. Reiman, J. M., Knutson, K. L. & Radisky, D. C. Immune promotion of epithelial-mesenchymal transition and generation of breast cancer stem cells. *Cancer Res* **70**, 3005–3008 (2010).

127. Gasic, G. J., Gasic, T. B., Galanti, N., Johnson, T. & Murphy, S. Platelet—tumor-cell interactions in mice. The role of platelets in the spread of malignant disease. *International Journal of Cancer* **11**, 704–718 (1973).
128. Milsom, C. *et al.* Tissue Factor and Cancer Stem Cells. *Arteriosclerosis, Thrombosis, and Vascular Biology* **29**, 2005 –2014 (2009).
129. Palumbo, J. S. *et al.* Fibrinogen is an important determinant of the metastatic potential of circulating tumor cells. *Blood* **96**, 3302 –3309 (2000).
130. Palumbo, J. S. *et al.* Platelets and fibrin(ogen) increase metastatic potential by impeding natural killer cell–mediated elimination of tumor cells. *Blood* **105**, 178 –185 (2005).
131. Nieswandt, B., Hafner, M., Echtenacher, B. & Männel, D. N. Lysis of Tumor Cells by Natural Killer Cells in Mice Is Impeded by Platelets. *Cancer Research* **59**, 1295 –1300 (1999).
132. Egan, K. *et al.* Platelet adhesion and degranulation induce pro-survival and pro-angiogenic signalling in ovarian cancer cells. *PLoS ONE* **6**, e26125 (2011).
133. Gay, L. J. & Felding-Habermann, B. Contribution of platelets to tumour metastasis. *Nat Rev Cancer* **11**, 123–134 (2011).
134. Oishi, N. & Wang, X. W. Novel therapeutic Strategies for Targeting Liver Cancer Stem Cells. *Int J Biol Sci* **7**, 517–535
135. Smith, L. M. *et al.* CD133/prominin-1 is a potential therapeutic target for antibody-drug conjugates in hepatocellular and gastric cancers. *Br. J. Cancer* **99**, 100–109 (2008).
136. Luu, H. H. *et al.* Wnt/beta-catenin signaling pathway as a novel cancer drug target. *Curr Cancer Drug Targets* **4**, 653–671 (2004).
137. Wei, W., Chua, M.-S., Grepper, S. & So, S. K. Soluble Frizzled-7 receptor inhibits Wnt signaling and sensitizes hepatocellular carcinoma cells towards doxorubicin. *Mol. Cancer* **10**, 16 (2011).
138. Zeng, G., Apte, U., Cieply, B., Singh, S. & Monga, S. P. S. siRNA-mediated beta-catenin knockdown in human hepatoma cells results in decreased growth and survival. *Neoplasia* **9**, 951–959 (2007).
139. Wei, W., Chua, M.-S., Grepper, S. & So, S. K. Blockade of Wnt-1 signaling leads to anti-tumor effects in hepatocellular carcinoma cells. *Mol. Cancer* **8**, 76 (2009).
140. Wei, W., Chua, M.-S., Grepper, S. & So, S. Small molecule antagonists of Tcf4/beta-catenin complex inhibit the growth of HCC cells in vitro and in vivo. *Int. J. Cancer* **126**, 2426–2436 (2010).
141. Prasad, C. P., Rath, G., Mathur, S., Bhatnagar, D. & Ralhan, R. Potent growth suppressive activity of curcumin in human breast cancer cells: Modulation of Wnt/beta-catenin signaling. *Chem. Biol. Interact.* **181**, 263–271 (2009).
142. Prosperi, J. R. & Goss, K. H. A Wnt-ow of opportunity: targeting the Wnt/beta-catenin pathway in breast cancer. *Curr Drug Targets* **11**, 1074–1088 (2010).
143. Huang, S.-M. A. *et al.* Tankyrase inhibition stabilizes axin and antagonizes Wnt signalling. *Nature* **461**, 614–620 (2009).
144. González-Sancho, J. M. *et al.* The Wnt antagonist DICKKOPF-1 gene is a downstream target of beta-catenin/TCF and is downregulated in human colon cancer. *Oncogene* **24**, 1098–1103 (2005).
145. Agur, Z. *et al.* Dickkopf1 regulates fate decision and drives breast cancer stem cells to differentiation: an experimentally supported mathematical model. *PLoS ONE* **6**, e24225 (2011).
146. Wang, Q. *et al.* Down-regulation of Sonic hedgehog signaling pathway activity is involved in 5-fluorouracil-induced apoptosis and motility inhibition in Hep3B cells. *Acta Biochim. Biophys. Sin. (Shanghai)* **40**, 819–829 (2008).
147. Lonardo, E. *et al.* Nodal/Activin signaling drives self-renewal and tumorigenicity of pancreatic cancer stem cells and provides a target for combined drug therapy. *Cell Stem Cell* **9**, 433–446 (2011).
148. Paul, S. & Dey, A. Wnt signaling and cancer development: therapeutic implication. *Neoplasma* **55**, 165–176 (2008).

149. Alison, M. R., Lim, S. M. & Nicholson, L. J. Cancer stem cells: problems for therapy? *The Journal of Pathology* **223**, 148–162 (2011).
150. Soprano, D. R., Teets, B. W. & Soprano, K. J. Role of retinoic acid in the differentiation of embryonal carcinoma and embryonic stem cells. *Vitam. Horm* **75**, 69–95 (2007).
151. Karsy, M., Albert, L., Tobias, M. E., Murali, R. & Jhanwar-Uniyal, M. All-trans retinoic acid modulates cancer stem cells of glioblastoma multiforme in an MAPK-dependent manner. *Anticancer Res.* **30**, 4915–4920 (2010).
152. Ying, M. *et al.* Regulation of glioblastoma stem cells by retinoic acid: role for Notch pathway inhibition. *Oncogene* **30**, 3454–3467 (2011).
153. Ginestier, C. *et al.* Retinoid signaling regulates breast cancer stem cell differentiation. *Cell Cycle* **8**, 3297–3302 (2009).
154. Kusmartsev, S. *et al.* All-trans-Retinoic Acid Eliminates Immature Myeloid Cells from Tumor-bearing Mice and Improves the Effect of Vaccination. *Cancer Research* **63**, 4441 –4449 (2003).
155. Mirza, N. *et al.* All-trans-Retinoic Acid Improves Differentiation of Myeloid Cells and Immune Response in Cancer Patients. *Cancer Research* **66**, 9299 –9307 (2006).
156. Gupta, P. B., Chaffer, C. L. & Weinberg, R. A. Cancer stem cells: mirage or reality? *Nat. Med.* **15**, 1010–1012 (2009).
157. Shackleton, M., Quintana, E., Fearon, E. R. & Morrison, S. J. Heterogeneity in Cancer: Cancer Stem Cells versus Clonal Evolution. *Cell* **138**, 822–829 (2009).
158. Li, F., Tiede, B., Massague, J. & Kang, Y. Beyond tumorigenesis: cancer stem cells in metastasis. *Cell Res* **17**, 3–14 (2007).
159. Dirks, P. Cancer stem cells: Invitation to a second round. *Nature* **466**, 40–41 (2010).
160. Roesch, A. *et al.* A temporarily distinct subpopulation of slow-cycling melanoma cells is required for continuous tumor growth. *Cell* **141**, 583–594 (2010).
161. Boiko, A. D. *et al.* Human melanoma-initiating cells express neural crest nerve growth factor receptor CD271. *Nature* **466**, 133–137 (2010).
162. Chaffer, C. L. *et al.* Normal and neoplastic nonstem cells can spontaneously convert to a stem-like state. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 7950–7955 (2011).
163. Quintana, E. *et al.* Efficient tumour formation by single human melanoma cells. *Nature* **456**, 593–598 (2008).
164. Quintana, E. *et al.* Phenotypic heterogeneity among tumorigenic melanoma cells from patients that is reversible and not hierarchically organized. *Cancer Cell* **18**, 510–523 (2010).
165. Nagler, C., Zänker, K. S. & Dittmar, T. Cell Fusion, Drug Resistance and Recurrence CSCs. *Adv. Exp. Med. Biol.* **714**, 173–182 (2011).
166. Lee, R. C., Feinbaum, R. L. & Ambros, V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**, 843–854 (1993).
167. Pasquinelli, A. E. *et al.* Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* **408**, 86–89 (2000).
168. Reinhart, B. J. *et al.* The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* **403**, 901–906 (2000).
169. miRBase. at <http://www.mirbase.org/index.shtml>
170. Lagos-Quintana, M. *et al.* Identification of tissue-specific microRNAs from mouse. *Curr. Biol.* **12**, 735–739 (2002).
171. Kim, V. N., Han, J. & Siomi, M. C. Biogenesis of small RNAs in animals. *Nat Rev Mol Cell Biol* **10**, 126–139 (2009).
172. Faller, M. & Guo, F. MicroRNA biogenesis: there's more than one way to skin a cat. *Biochim Biophys Acta* **1779**, 663–667 (2008).
173. Czech, B. & Hannon, G. J. Small RNA sorting: matchmaking for Argonautes. *Nat Rev Genet* **12**, 19–31 (2011).
174. Okamura, K., Hagen, J. W., Duan, H., Tyler, D. M. & Lai, E. C. The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. *Cell* **130**, 89–100 (2007).

175. Morin, R. D. *et al.* Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Res* **18**, 610–621 (2008).
176. Pratt, A. J. & MacRae, I. J. The RNA-induced Silencing Complex: A Versatile Gene-silencing Machine. *Journal of Biological Chemistry* **284**, 17897 –17901 (2009).
177. Wang, B. *et al.* Distinct passenger strand and mRNA cleavage activities of human Argonaute proteins. *Nat Struct Mol Biol* **16**, 1259–1266 (2009).
178. Lim, L. P. *et al.* Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**, 769–773 (2005).
179. Krol, J., Loedige, I. & Filipowicz, W. The widespread regulation of microRNA biogenesis, function and decay. *Nature Reviews Genetics* (2010).doi:10.1038/nrg2843
180. Gu, S. *et al.* Thermodynamic stability of small hairpin RNAs highly influences the loading process of different mammalian Argonautes. *Proceedings of the National Academy of Sciences* **108**, 9208 –9213 (2011).
181. Kawamata, T., Seitz, H. & Tomari, Y. Structural determinants of miRNAs for RISC loading and slicer-independent unwinding. *Nat Struct Mol Biol* **16**, 953–960 (2009).
182. Jazdzewski, K. *et al.* Polymorphic mature microRNAs from passenger strand of pre-miR-146a contribute to thyroid cancer. *Proceedings of the National Academy of Sciences* **106**, 1502 – 1505 (2009).
183. Kim, S. *et al.* MicroRNA miR-199a* regulates the MET proto-oncogene and the downstream extracellular signal-regulated kinase 2 (ERK2). *J. Biol. Chem.* **283**, 18158–18166 (2008).
184. Mah, S. M., Buske, C., Humphries, R. K. & Kuchenbauer, F. miRNA*: a passenger stranded in RNA-induced silencing complex? *Crit. Rev. Eukaryot. Gene Expr.* **20**, 141–148 (2010).
185. Kawamata, T. & Tomari, Y. Making RISC. *Trends in Biochemical Sciences* **35**, 368–376 (2010).
186. Liu, J. *et al.* A role for the P-body component GW182 in microRNA function. *Nat Cell Biol* **7**, 1261–1266 (2005).
187. Fabian, M. R. *et al.* miRNA-mediated deadenylation is orchestrated by GW182 through two conserved motifs that interact with CCR4-NOT. *Nat. Struct. Mol. Biol.* **18**, 1211–1217 (2011).
188. Filipowicz, W., Bhattacharyya, S. N. & Sonenberg, N. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat Rev Genet* **9**, 102–114 (2008).
189. Tang, R. & Zen, K. Gold glitters everywhere: nucleus microRNAs and their functions. *Frontiers in Biology* **6**, 69–75 (2011).
190. Tan, Y. *et al.* Transcriptional inhibiton of Hoxd4 expression by miRNA-10a in human breast cancer cells. *BMC Mol Biol* **10**, 12–12
191. Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15–20 (2005).
192. Friedman, R. C., Farh, K. K.-H., Burge, C. B. & Bartel, D. P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Research* **19**, 92–105 (2009).
193. Peter, M. E. Targeting of mRNAs by multiple miRNAs: the next step. *Oncogene* **29**, 2161–2164 (2010).
194. Wu, S. *et al.* Multiple microRNAs modulate p21Cip1/Waf1 expression by directly targeting its 3' untranslated region. *Oncogene* **29**, 2302–2308 (2010).
195. Doench, J. G. & Sharp, P. A. Specificity of microRNA target selection in translational repression. *Genes Dev* **18**, 504–11 (2004).
196. Brodersen, P. & Voinnet, O. Revisiting the principles of microRNA target recognition and mode of action. *Nat Rev Mol Cell Biol* **10**, 141–148 (2009).
197. Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U. & Segal, E. The role of site accessibility in microRNA target recognition. *Nat Genet* **39**, 1278–1284 (2007).
198. Hofacker, I. L. How microRNAs choose their targets. *Nat Genet* **39**, 1191–1192 (2007).
199. Tay, Y., Zhang, J., Thomson, A. M., Lim, B. & Rigoutsos, I. MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature* **455**, 1124–8 (2008).
200. Zhou, X., Duan, X., Qian, J. & Li, F. Abundant conserved microRNA target sites in the 5'-untranslated region and coding sequence. *Genetica* **137**, 159–164 (2009).

201. Lee, I. *et al.* New class of microRNA targets containing simultaneous 5'-UTR and 3'-UTR interaction sites. *Genome Res.* **19**, 1175–1183 (2009).
202. Flynt, A. S. & Lai, E. C. Biological principles of microRNA-mediated regulation: shared themes amid diversity. *Nat Rev Genet* **9**, 831–842 (2008).
203. Huang, Y. & Gu, X. A study of the evolution of human microRNAs by their apparent repression effectiveness on target genes. *PLoS ONE* **6**, e25034 (2011).
204. Lutter, D., Marr, C., Krumsiek, J., Lang, E. W. & Theis, F. J. Intronic microRNAs support their host genes by mediating synergistic and antagonistic regulatory effects. *BMC Genomics* **11**, 224 (2010).
205. Hollstein, M., Sidransky, D., Vogelstein, B. & Harris, C. C. p53 mutations in human cancers. *Science* **253**, 49–53 (1991).
206. Chang, C.-J. *et al.* p53 regulates epithelial-mesenchymal transition and stem cell properties through modulating miRNAs. *Nat Cell Biol* **13**, 317–323 (2011).
207. Xu, N., Papagiannakopoulos, T., Pan, G., Thomson, J. A. & Kosik, K. S. MicroRNA-145 Regulates OCT4, SOX2, and KLF4 and Represses Pluripotency in Human Embryonic Stem Cells. *Cell* **137**, 647–658 (2009).
208. Saito, Y. *et al.* Specific activation of microRNA-127 with downregulation of the proto-oncogene BCL6 by chromatin-modifying drugs in human cancer cells. *Cancer Cell* **9**, 435–443 (2006).
209. Han, J. *et al.* Posttranscriptional crossregulation between Drosha and DGCR8. *Cell* **136**, 75–84 (2009).
210. Ebert, M. S. & Sharp, P. A. Emerging roles for natural microRNA sponges. *Curr. Biol.* **20**, R858–861 (2010).
211. Yu, J. *et al.* MicroRNA-184 antagonizes microRNA-205 to maintain SHIP2 levels in epithelia. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 19300–19305 (2008).
212. Place, R. F., Li, L.-C., Pookot, D., Noonan, E. J. & Dahiya, R. MicroRNA-373 induces expression of genes with complementary promoter sequences. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 1608–1613 (2008).
213. Ørom, U. A., Nielsen, F. C. & Lund, A. H. MicroRNA-10a Binds the 5'UTR of Ribosomal Protein mRNAs and Enhances Their Translation. *Molecular Cell* **30**, 460–471 (2008).
214. Valadi, H. *et al.* Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nat Cell Biol* **9**, 654–659 (2007).
215. Calin, G. A. *et al.* Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A* **99**, 15524–15529 (2002).
216. Calin, G. A. *et al.* Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 2999 –3004 (2004).
217. Lu, J. *et al.* MicroRNA expression profiles classify human cancers. *Nature* **435**, 834–838 (2005).
218. Iorio, M. V. *et al.* MicroRNA gene expression deregulation in human breast cancer. *Cancer Res.* **65**, 7065–7070 (2005).
219. Garzon, R., Calin, G. A. & Croce, C. M. MicroRNAs in Cancer. *Annual Review of Medicine* **60**, 167–179 (2009).
220. Si, M.-L. *et al.* miR-21-mediated tumor growth. *Oncogene* **26**, 2799–2803 (2006).
221. Klein, U. *et al.* The DLEU2/miR-15a/16-1 Cluster Controls B Cell Proliferation and Its Deletion Leads to Chronic Lymphocytic Leukemia. *Cancer Cell* **17**, 28–40 (2010).
222. Tao, J. *et al.* microRNA-21 modulates cell proliferation and sensitivity to doxorubicin in bladder cancer cells. *Oncol. Rep.* **25**, 1721–1729 (2011).
223. Yao, Q., Xu, H., Zhang, Q.-Q., Zhou, H. & Qu, L.-H. MicroRNA-21 promotes cell proliferation and down-regulates the expression of programmed cell death 4 (PDCD4) in HeLa cervical carcinoma cells. *Biochem. Biophys. Res. Commun.* **388**, 539–542 (2009).

224. Chan, J. A., Krichevsky, A. M. & Kosik, K. S. MicroRNA-21 is an antiapoptotic factor in human glioblastoma cells. *Cancer Res.* **65**, 6029–6033 (2005).
225. Diederichs, S. & Haber, D. A. Sequence variations of microRNAs in human cancer: alterations in predicted secondary structure do not affect processing. *Cancer Res.* **66**, 6097–6104 (2006).
226. Roldo, C. *et al.* MicroRNA expression abnormalities in pancreatic endocrine and acinar tumors are associated with distinctive pathologic features and clinical behavior. *J. Clin. Oncol.* **24**, 4677–4684 (2006).
227. Volinia, S. *et al.* A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 2257–2261 (2006).
228. Lawrie, C. H. *et al.* MicroRNA expression distinguishes between germinal center B cell-like and activated B cell-like subtypes of diffuse large B cell lymphoma. *Int. J. Cancer* **121**, 1156–1161 (2007).
229. Calin, G. A. *et al.* A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *N. Engl. J. Med.* **353**, 1793–1801 (2005).
230. Meng, F. *et al.* MicroRNA-21 regulates expression of the PTEN tumor suppressor gene in human hepatocellular cancer. *Gastroenterology* **133**, 647–658 (2007).
231. Grignol, V. *et al.* miR-21 and miR-155 are associated with mitotic activity and lesion depth of borderline melanocytic lesions. *Br. J. Cancer* **105**, 1023–1029 (2011).
232. Ahmed, M. I., Mardaryev, A. N., Lewis, C. J., Sharov, A. A. & Botchkareva, N. V. MicroRNA-21 is an important downstream component of BMP signalling in epidermal keratinocytes. *J. Cell. Sci.* **124**, 3399–3404 (2011).
233. Lu, Z. *et al.* MicroRNA-21 promotes cell transformation by targeting the programmed cell death 4 gene. *Oncogene* **27**, 4373–4379 (2008).
234. Yu, Y. *et al.* MicroRNA-21 induces stemness by downregulating transforming growth factor beta receptor 2 (TGFβR2) in colon cancer cells. *Carcinogenesis* (2011).doi:10.1093/carcin/bgr246
235. Chang, T.-C. *et al.* Widespread microRNA repression by Myc contributes to tumorigenesis. *Nat. Genet.* **40**, 43–50 (2008).
236. Weber, B., Stresemann, C., Brueckner, B. & Lyko, F. Methylation of Human MicroRNA Genes in Normal and Neoplastic Cells. *Cell Cycle* **6**, 1001–1005 (2007).
237. Tsuruta, T. *et al.* miR-152 Is a Tumor Suppressor microRNA That Is Silenced by DNA Hypermethylation in Endometrial Cancer. *Cancer Res.* **71**, 6450–6462 (2011).
238. Brueckner, B. *et al.* The Human let-7a-3 Locus Contains an Epigenetically Regulated MicroRNA Gene with Oncogenic Function. *Cancer Research* **67**, 1419 –1423 (2007).
239. Cheung, H.-H. *et al.* Methylation of an intronic region regulates miR-199a in testicular tumor malignancy. *Oncogene* **30**, 3404–3415 (2011).
240. Izumiya, M., Tsuchiya, N., Okamoto, K. & Nakagama, H. Systematic exploration of cancer-associated microRNA through functional screening assays. *Cancer Science* **102**, 1615–1621 (2011).
241. Chia, N.-Y. *et al.* A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature* **468**, 316–320 (2010).
242. Kosaka, N., Iguchi, H. & Ochiya, T. Circulating microRNA in body fluid: a new potential biomarker for cancer diagnosis and prognosis. *Cancer Science* **101**, 2087–2092 (2010).
243. Chapin, S. C. & Doyle, P. S. Ultrasensitive multiplexed microRNA quantification on encoded gel microparticles using rolling circle amplification. *Anal. Chem.* **83**, 7179–7185 (2011).
244. Garzon, R., Marcucci, G. & Croce, C. M. Targeting microRNAs in cancer: rationale, strategies and challenges. *Nat Rev Drug Discov* **9**, 775–789 (2010).
245. Wang, Z. *et al.* Targeting miRNAs involved in cancer stem cell and EMT regulation: an emerging concept in overcoming drug resistance. *Drug Resist Updat* **13**, 109–118 (2010).
246. Bader, A. G., Brown, D., Stoudemire, J. & Lammers, P. Developing therapeutic microRNAs for cancer. *Gene Ther* (2011).at <http://dx.doi.org/10.1038/gt.2011.79>

247. Zhang, S. *et al.* Aryl Hydrocarbon Receptor (AHR) Agonists Induce MicroRNA-335 Expression and Inhibit Lung Metastasis of Estrogen Receptor Negative Breast Cancer Cells. *Molecular Cancer Therapeutics* (2011).doi:10.1158/1535-7163.MCT-11-0548
248. Garzon, R. *et al.* MicroRNA gene expression during retinoic acid-induced differentiation of human acute promyelocytic leukemia. *Oncogene* **26**, 4148–4157 (2007).
249. Kota, J. *et al.* Therapeutic microRNA delivery suppresses tumorigenesis in a murine liver cancer model. *Cell* **137**, 1005–1017 (2009).
250. Gangaraju, V. K. & Lin, H. MicroRNAs: key regulators of stem cells. *Nat Rev Mol Cell Biol* **10**, 116–125 (2009).
251. Wang, Y. *et al.* Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation. *Nat. Genet.* **40**, 1478–1483 (2008).
252. Murchison, E. P., Partridge, J. F., Tam, O. H., Cheloufi, S. & Hannon, G. J. Characterization of Dicer-deficient murine embryonic stem cells. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 12135–12140 (2005).
253. Strauss, W. M., Chen, C., Lee, C.-T. & Ridzon, D. Nonrestrictive developmental regulation of microRNA gene expression. *Mamm. Genome* **17**, 833–840 (2006).
254. Shimono, Y. *et al.* Downregulation of miRNA-200c links breast cancer stem cells with normal stem cells. *Cell* **138**, 592–603 (2009).
255. Wellner, U. *et al.* The EMT-activator ZEB1 promotes tumorigenicity by repressing stemness-inhibiting microRNAs. *Nat. Cell Biol.* **11**, 1487–1495 (2009).
256. Yu, F. *et al.* let-7 regulates self renewal and tumorigenicity of breast cancer cells. *Cell* **131**, 1109–1123 (2007).
257. McDermott, S. P. & Wicha, M. S. Targeting Breast Cancer Stem Cells. *Mol Oncol* **4**, 404–419 (2010).
258. Liu, C. & Tang, D. G. MicroRNA Regulation of Cancer Stem Cells. *Cancer Research* **71**, 5950 – 5954 (2011).
259. Büssing, I., Slack, F. J. & Grosshans, H. let-7 microRNAs in development, stem cells and cancer. *Trends Mol Med* **14**, 400–409 (2008).
260. van Haaften, G. & Agami, R. Tumorigenicity of the miR-17-92 cluster distilled. *Genes & Development* **24**, 1 –4 (2010).
261. Monzo, M. *et al.* Overlapping expression of microRNAs in human embryonic colon and colorectal cancer. *Cell Res* **18**, 823–833 (2008).
262. He, L. *et al.* A microRNA polycistron as a potential human oncogene. *Nature* **435**, 828–833 (2005).
263. Wong, P. *et al.* The miR-17-92 microRNA Polycistron Regulates MLL Leukemia Stem Cell Potential by Modulating p21 Expression. *Cancer Res* **70**, 3833–3842 (2010).
264. Hohjoh, H. & Fukushima, T. Marked change in microRNA expression during neuronal differentiation of human teratocarcinoma NTera2D1 and mouse embryonal carcinoma P19 cells. *Biochem. Biophys. Res. Commun.* **362**, 360–367 (2007).
265. Lee, N. S. *et al.* miR-302b maintains 'stemness' of human embryonal carcinoma cells by post-transcriptional regulation of Cyclin D2 expression. *Biochemical and Biophysical Research Communications* **377**, 434–440 (2008).
266. Greer Card, D. A. *et al.* Oct4/Sox2-Regulated miR-302 Targets Cyclin D1 in Human Embryonic Stem Cells. *Mol Cell Biol* **28**, 6426–6438 (2008).
267. Garzia, L. *et al.* MicroRNA-199b-5p Impairs Cancer Stem Cells through Negative Regulation of HES1 in Medulloblastoma. *PLoS ONE* **4**, e4998 (2009).
268. Zhang, J. *et al.* microRNA-150 inhibits human CD133-positive liver cancer stem cells through negative regulation of the transcription factor c-Myb. *International Journal of Oncology* (2011).doi:10.3892/ijo.2011.1242
269. Nalls, D., Tang, S.-N., Rodova, M., Srivastava, R. K. & Shankar, S. Targeting epigenetic regulation of miR-34a for treatment of pancreatic cancer by inhibition of pancreatic cancer stem cells. *PLoS ONE* **6**, e24099 (2011).

270. Schraivogel, D. *et al.* CAMTA1 is a novel tumour suppressor regulated by miR-9/9* in glioblastoma stem cells. *EMBO J.* **30**, 4309–4322 (2011).
271. Ji, J. *et al.* Identification of microRNA-181 by genome-wide screening as a critical player in EpCAM-positive hepatic cancer stem cells. *Hepatology* **50**, 472–480 (2009).
272. Bitarte, N. *et al.* MicroRNA-451 Is Involved in the Self-renewal, Tumorigenicity, and Chemoresistance of Colorectal Cancer Stem Cells. *Stem Cells* **29**, 1661–1671 (2011).
273. DeSano, J. T. & Xu, L. MicroRNA Regulation of Cancer Stem Cells and Therapeutic Implications. *AAPS J* **11**, 682–692
274. Gallagher, M. F. *et al.* Regulation of microRNA biosynthesis and expression in 2102Ep embryonal carcinoma stem cells is mirrored in ovarian serous adenocarcinoma patients. *J Ovarian Res* **2**, 19 (2009).
275. Vêncio, E. F. *et al.* Embryonal carcinoma cell induction of miRNA and mRNA changes in co-cultured prostate stromal fibromuscular cells. *J. Cell. Physiol.* **226**, 1479–1488 (2011).
276. Liu, H. *et al.* Oct4 regulates the miR-302 cluster in P19 mouse embryonic carcinoma cells. *Mol. Biol. Rep.* **38**, 2155–2160 (2011).
277. Koster, R. *et al.* Cytoplasmic p21 expression levels determine cisplatin resistance in human testicular cancer. *J. Clin. Invest.* **120**, 3594–3605 (2010).
278. Lian, J. *et al.* Downregulation of microRNA-383 is associated with male infertility and promotes testicular embryonal carcinoma cell proliferation by targeting IRF1. *Cell Death Dis* **1**, e94– (2010).
279. Matthaei, K. I., Andrews, P. W. & Bronson, D. L. Retinoic acid fails to induce differentiation in human teratocarcinoma cell lines that express high levels of a cellular receptor protein. *Exp. Cell Res.* **143**, 471–474 (1983).
280. Josephson, R. *et al.* Qualification of embryonal carcinoma 2102Ep as a reference for human embryonic stem cell research. *Stem Cells* **25**, 437–446 (2007).
281. Andrews, P. W., Fenderson, B. & Hakomori, S. Human embryonal carcinoma cells and their differentiation in culture. *International Journal of Andrology* **10**, 95–104 (1987).
282. Clark, A. T. The stem cell identity of testicular cancer. *Stem Cell Rev* **3**, 49–59 (2007).
283. Hovatta, O. *et al.* A Teratocarcinoma-Like Human Embryonic Stem Cell (hESC) Line and Four hESC Lines Reveal Potentially Oncogenic Genomic Changes. *PLoS ONE* **5**, e10263 (2010).
284. Benham, F. J., Andrews, P. W., Knowles, B. B., Bronson, D. L. & Harris, H. Alkaline phosphatase isozymes as possible markers of differentiation in human testicular teratocarcinoma cell lines. *Developmental Biology* **88**, 279–287 (1981).
285. Ryan, J. A. *Understanding and managing cell culture contamination*. (Corning Incorporated: 1994).
286. Eda, A., Tamura, Y., Yoshida, M. & Hohjoh, H. Systematic gene regulation involving miRNAs during neuronal differentiation of mouse P19 embryonic carcinoma cell. *Biochemical and Biophysical Research Communications* **388**, 648–653 (2009).
287. Horwich, M. D. & Zamore, P. D. Design and delivery of antisense oligonucleotides to block microRNA function in cultured Drosophila and human cells. *Nat Protoc* **3**, 1537–1549 (2008).
288. Hawkins, P. G., Santoso, S., Adams, C., Anest, V. & Morris, K. V. Promoter targeted small RNAs induce long-term transcriptional gene silencing in human cells. *Nucleic Acids Res.* **37**, 2984–2995 (2009).
289. Hentschke, M., Kurth, I., Borgmeyer, U. & Hubner, C. A. Germ Cell Nuclear Factor Is a Repressor of CRIPTO-1 and CRIPTO-3. *Journal of Biological Chemistry* **281**, 33497–33504 (2006).
290. You, J. S. *et al.* Depletion of Embryonic Stem Cell Signature by Histone Deacetylase Inhibitor in NCCIT Cells: Involvement of Nanog Suppression. *Cancer Research* **69**, 5716–5725 (2009).
291. Benito, A. *et al.* A novel role for par bZIP proteins in the transcriptional regulation of a BH3-only proapoptotic gene. *Journal of Biological Chemistry* (2006).
292. Böhm, J., Sustmann, C., Wilhelm, C. & Kohlhase, J. SALL4 is directly activated by TCF/LEF in the canonical Wnt signaling pathway. *Biochemical and Biophysical Research Communications* **348**, 898–907 (2006).

293. Lipofectamine® RNAiMAX Transfection Reagent. at <http://www.invitrogen.com/site/us/en/home/Products-and-Services/Applications/Cell-Culture/Transfection/Transfection___Selection-Misc/lipofectamine-rnaimax-transfection-reagent.html>
294. Fujita, S. *et al.* Highly efficient reverse transfection with siRNA in multiple wells of microtiter plates. *Journal of Bioscience and Bioengineering* **104**, 329–333 (2007).
295. Amarzguioui, M. Improved siRNA-mediated silencing in refractory adherent cell lines by detachment and transfection in suspension. *BioTechniques* **36**, 766–768, 770 (2004).
296. Duran, C. *et al.* Hybrids of pluripotent and nullipotent human embryonal carcinoma cells: Partial retention of a pluripotent phenotype. *International Journal of Cancer* **93**, 324–332 (2001).
297. Matin, M. M. *et al.* Specific Knockdown of Oct4 and β2-microglobulin Expression by RNA Interference in Human Embryonic Stem Cells and Embryonic Carcinoma Cells. *Stem Cells* **22**, 659–668 (2004).
298. Wang, S., Rosengren, L. E., Franlund, M., Hamberger, A. & Haglid, K. G. Bcl-2 expression regulates cell sensitivity to S100[beta]-mediated apoptosis. *Molecular Brain Research* **70**, 167–176 (1999).
299. Kozomara, A. & Griffiths-Jones, S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* **39**, D152–157 (2011).
300. Okoniewski, M. J., Hey, Y., Pepper, S. D. & Miller, C. J. High correspondence between Affymetrix exon and standard expression arrays. *BioTechniques* **42**, 181–185 (2007).
301. Haan, C. & Behrmann, I. A cost effective non-commercial ECL-solution for Western blot detections yielding strong signals and low background. *Journal of immunological methods* **318**, 11–19 (2007).
302. Reya, T., Morrison, S. J., Clarke, M. F. & Weissman, I. L. Stem cells, cancer, and cancer stem cells. *Nature* **414**, 105–111 (2001).
303. Pece, S. *et al.* Biological and molecular heterogeneity of breast cancers correlates with their cancer stem cell content. *Cell* **140**, 62–73 (2010).
304. Dylla, S. J. *et al.* Colorectal Cancer Stem Cells Are Enriched in Xenogeneic Tumors Following Chemotherapy. *PLoS ONE* **3**, e2428 (2008).
305. DiMasi, J. A. & Grabowski, H. G. Economics of New Oncology Drug Development. *Journal of Clinical Oncology* **25**, 209 –216 (2007).
306. Mina, L. A. & Sledge, G. W. Rethinking the metastatic cascade as a therapeutic target. *Nat Rev Clin Oncol* **8**, 325–332 (2011).
307. Dingli, D. & Michor, F. Successful Therapy Must Eradicate Cancer Stem Cells. *STEM CELLS* **24**, 2603–2610 (2006).
308. Freemantle, S. J., Dragnev, K. H. & Dmitrovsky, E. The Retinoic Acid Paradox in Cancer Chemoprevention. *Journal of the National Cancer Institute* **98**, 426 –427 (2006).
309. Jerome J., K. Toxicology, carcinogenicity, and teratogenicity of some orally administered retinoids. *Journal of the American Academy of Dermatology* **6**, 652–659 (1982).
310. Sulik, K. K., Dehart, D. B., Rogers, J. M. & Chernoff, N. Teratogenicity of low doses of all-trans retinoic acid in presomite mouse embryos. *Teratology* **51**, 398–403 (1995).
311. Rothman, K. J. *et al.* Teratogenicity of high vitamin A intake. *N. Engl. J. Med.* **333**, 1369–1373 (1995).
312. Chen, Y., Zhu, X., Zhang, X., Liu, B. & Huang, L. Nanoparticles Modified With Tumor-targeting scFv Deliver siRNA and miRNA for Cancer Therapy. *Mol Ther* (2010).doi:10.1038/mt.2010.136
313. Andrews, P. W. Human embryonal carcinoma cells in culture do not synthesize fibronectin until they differentiate. *Int. J. Cancer* **30**, 567–571 (1982).
314. Andrews, P. W. Retinoic acid induces neuronal differentiation of a cloned human embryonal carcinoma cell line in vitro. *Dev. Biol.* **103**, 285–293 (1984).
315. Lee, V. M. & Andrews, P. W. Differentiation of NTERA-2 clonal human embryonal carcinoma cells into neurons involves the induction of all three neurofilament proteins. *J. Neurosci.* **6**, 514–521 (1986).

316. Fenderson, B. A., Andrews, P. W., Nudelman, E., Clausen, H. & Hakomori, S. Glycolipid core structure switching from globo- to lacto- and ganglio-series during retinoic acid-induced differentiation of TERA-2-derived human embryonal carcinoma cells. *Dev. Biol.* **122**, 21–34 (1987).
317. Pleasure, S. J. & Lee, V. M. -Y NTera 2 Cells: A human cell line which displays characteristics expected of a human committed neuronal progenitor cell. *Journal of Neuroscience Research* **35**, 585–602 (1993).
318. Schwartz, C. M. *et al.* NTera2: a model system to study dopaminergic differentiation of human embryonic stem cells. *Stem Cells and Development* **14**, 517–534 (2005).
319. Thomas, M., Lieberman, J. & Lal, A. Desperately seeking microRNA targets. *Nat Struct Mol Biol* **17**, 1169–1174 (2010).
320. Krek, A. *et al.* Combinatorial microRNA target predictions. *Nat. Genet.* **37**, 495–500 (2005).
321. Kiriakidou, M. *et al.* A combined computational-experimental approach predicts human microRNA targets. *Genes & Development* **18**, 1165–1178 (2004).
322. Betel, D., Wilson, M., Gabow, A., Marks, D. S. & Sander, C. The microRNA.org resource: targets and expression. *Nucl. Acids Res.* **36**, D149–153 (2008).
323. Miranda, K. C. *et al.* A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell* **126**, 1203–1217 (2006).
324. Gennarino, V. A. *et al.* HOCTAR database: a unique resource for microRNA target prediction. *Gene* **480**, 51–58 (2011).
325. Gennarino, V. A. *et al.* MicroRNA target prediction by expression analysis of host genes. *Genome Res.* **19**, 481–490 (2009).
326. Hammell, M. *et al.* mirWIP: microRNA target prediction based on microRNA-containing ribonucleoprotein-enriched transcripts. *Nat. Methods* **5**, 813–819 (2008).
327. Yang, J.-H. *et al.* starBase: a database for exploring microRNA-mRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data. *Nucleic Acids Research* **39**, D202–D209 (2010).
328. LaMartina, S. & Krebs, J. Conveniently Monitor miRNA Transfection and Function at the Protein Level (Ambion TechNotes 14:2). at <http://www.ambion.com/techlib/tn/142/8.html>
329. Grimson, A. *et al.* MicroRNA Targeting Specificity in Mammals: Determinants beyond Seed Pairing. *Molecular Cell* **27**, 91–105 (2007).
330. Min, H. & Yoon, S. Got target?: computational methods for microRNA target prediction and their extension. *Exp Mol Med* **42**, 233–244 (2010).
331. Draper, N. R. *Applied Regression Analysis*. (Wiley: New York, 1966).
332. Kutner, M., Nachtsheim, C., Neter, J. & Li, W. *Applied Linear Statistical Models*. (McGraw-Hill/Irwin: 2004).
333. Xu, H. *et al.* WWP2 promotes degradation of transcription factor OCT4 in human embryonic stem cells. *Cell Res* **19**, 561–573 (2009).
334. Gilbert, D. C. *et al.* Clinical and biological significance of CXCL12 and CXCR4 expression in adult testes and germ cell tumours of adults and adolescents. *The Journal of Pathology* **217**, 94–102 (2009).
335. Cai, J. *et al.* Assessing self-renewal and differentiation in human embryonic stem cell lines. *Stem Cells* **24**, 516–530 (2006).
336. Esquela-Kerscher, A. & Slack, F. J. Oncomirs [mdash] microRNAs with a role in cancer. *Nat Rev Cancer* **6**, 259–269 (2006).
337. Wu, L. & Belasco, J. G. Micro-RNA Regulation of the Mammalian lin-28 Gene during Neuronal Differentiation of Embryonal Carcinoma Cells. *Mol. Cell. Biol.* **25**, 9198–9208 (2005).
338. Sempere, L. F. *et al.* Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation. *Genome Biol.* **5**, R13 (2004).
339. Gunaratne, P. H. Embryonic stem cell microRNAs: defining factors in induced pluripotent (iPS) and cancer (CSC) stem cells? *Curr Stem Cell Res Ther* **4**, 168–177 (2009).

340. Betel, D., Koppal, A., Agius, P., Sander, C. & Leslie, C. Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol.* **11**, R90 (2010).
341. John, B. *et al.* Human MicroRNA Targets. *PLoS Biol* **2**, e363 (2004).
342. Karaczyn, A. *et al.* Two novel human NUMB isoforms provide a potential link between development and cancer. *Neural Development* **5**, 31 (2010).
343. Pece, S., Confalonieri, S., R. Romano, P. & Di Fiore, P. P. NUMB-ing down cancer by more than just a NOTCH. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* **1815**, 26–43 (2011).
344. Agirre, X. *et al.* Down-regulation of hsa-miR-10a in chronic myeloid leukemia CD34+ cells increases USF2-mediated cell growth. *Mol. Cancer Res.* **6**, 1830–1840 (2008).
345. Malzkorn, B. *et al.* Identification and Functional Characterization of microRNAs Involved in the Malignant Progression of Gliomas. *Brain Pathol* (2009).doi:10.1111/j.1750-3639.2009.00328.x
346. Ovcharenko, D., Kelnar, K., Johnson, C., Leng, N. & Brown, D. Genome-scale microRNA and small interfering RNA screens identify small RNA modulators of TRAIL-induced apoptosis pathway. *Cancer research* **67**, 10782 (2007).
347. Park, C. W., Zeng, Y., Zhang, X., Subramanian, S. & Steer, C. J. Mature microRNAs identified in highly purified nuclei from HCT116 colon cancer cells. *RNA biology* **7**, 606 (2010).
348. Tzur, G. *et al.* MicroRNA Expression Patterns and Function in Endodermal Differentiation of Human Embryonic Stem Cells. *PLoS ONE* **3**, e3726 (2008).
349. Wei, W. *et al.* Global Egr1-miRNAs Binding Analysis in PMA-Induced K562 Cells Using ChIP-Seq. *Journal of Biomedicine and Biotechnology* **2010**, (2010).
350. Weiss, F. U. *et al.* Retinoic Acid Receptor Antagonists Inhibit miR-10a Expression and Block Metastatic Behavior of Pancreatic Cancer. *Gastroenterology* **137**, 2136–2145.e7 (2009).
351. Ovcharenko, D. *et al.* miR-10a overexpression is associated with NPM1 mutations and MDM4 downregulation in intermediate-risk acute myeloid leukemia. *Exp. Hematol.* **39**, 1030–1042.e7 (2011).
352. Foley, N. H. *et al.* MicroRNAs 10a and 10b are potent inducers of neuroblastoma cell differentiation through targeting of nuclear receptor corepressor 2. *Cell Death Differ.* **18**, 1089–1098 (2011).
353. Meseguer, S., Mudduluru, G., Escamilla, J. M., Allgayer, H. & Barettino, D. MicroRNAs-10a and -10b contribute to retinoic acid-induced differentiation of neuroblastoma cells and target the alternative splicing regulatory factor SFRS1 (SF2/ASF). *J. Biol. Chem.* **286**, 4150–4164 (2011).
354. Huang, H. *et al.* miR-10a contributes to retinoid acid-induced smooth muscle cell differentiation. *J. Biol. Chem* **285**, 9383–9389 (2010).
355. Li, D. *et al.* MicroRNA-99a inhibits hepatocellular carcinoma growth and correlates with prognosis of hepatocellular carcinoma patients. *The Journal of Biological Chemistry* (2011).doi:10.1074/jbc.M111.270561
356. Port, M. *et al.* Micro-RNA expression in cisplatin resistant germ cell tumor cell lines. *Mol. Cancer* **10**, 52 (2011).
357. Oneyama, C. *et al.* MicroRNA-mediated downregulation of mTOR/FGFR3 controls tumor growth induced by Src-related oncogenic pathways. *Oncogene* **30**, 3489–3501 (2011).
358. Sun, D. *et al.* miR-99 family of MicroRNAs suppresses the expression of prostate-specific antigen and prostate cancer cell proliferation. *Cancer Res.* **71**, 1313–1324 (2011).
359. Candiani, S., Moronti, L., De Pietri Tonelli, D., Garbarino, G. & Pestarino, M. A study of neural-related microRNAs in the developing amphioxus. *Evodevo* **2**, 15 (2011).
360. Li, P. *et al.* Localized expression pattern of miR-184 in Drosophila. *Mol. Biol. Rep.* **38**, 355–358 (2011).
361. Iovino, N., Pane, A. & Gaul, U. miR-184 has multiple roles in Drosophila female germline development. *Dev. Cell* **17**, 123–133 (2009).
362. Foley, N. H. *et al.* MicroRNA-184 inhibits neuroblastoma cell survival through targeting the serine/threonine kinase AKT2. *Mol. Cancer* **9**, 83 (2010).

363. Tivnan, A., Foley, N. H., Tracey, L., Davidoff, A. M. & Stallings, R. L. MicroRNA-184-mediated inhibition of tumour growth in an orthotopic murine model of neuroblastoma. *Anticancer Res.* **30**, 4391–4395 (2010).
364. Liu, C. *et al.* Epigenetic regulation of miR-184 by MBD1 governs neural stem cell proliferation and differentiation. *Cell Stem Cell* **6**, 433–444 (2010).
365. Zhao, X. *et al.* MicroRNA-mediated control of oligodendrocyte differentiation. *Neuron* **65**, 612–626 (2010).
366. Dugas, J. C. *et al.* Dicer1 and miR-219 Are required for normal oligodendrocyte differentiation and myelination. *Neuron* **65**, 597–611 (2010).
367. Xu, Y. *et al.* MicroRNA-335 acts as a metastasis suppressor in gastric cancer by targeting Bcl-w and specificity protein 1. *Oncogene* (2011).doi:10.1038/onc.2011.340
368. Heyn, H. *et al.* MicroRNA miR-335 is crucial for the BRCA1 regulatory cascade in breast cancer development. *International Journal of Cancer. Journal International Du Cancer* (2011).doi:10.1002/ijc.25962
369. Shu, M. *et al.* Targeting oncogenic miR-335 inhibits growth and invasion of malignant astrocytoma cells. *Mol Cancer* **10**, 59 (2011).
370. Zhang, J. *et al.* Effects of miR-335-5p in modulating osteogenic differentiation by specifically downregulating Wnt antagonist DKK1. *J. Bone Miner. Res.* **26**, 1953–1963 (2011).
371. Png, K. J. *et al.* MicroRNA-335 inhibits tumor reinitiation and is silenced through genetic and epigenetic mechanisms in human breast cancer. *Genes Dev* **25**, 226–231 (2011).
372. Tomé, M. *et al.* miR-335 orchestrates cell proliferation, migration and differentiation in human mesenchymal stem cells. *Cell Death Differ* (2010).doi:10.1038/cdd.2010.167
373. Tavazoie, S. F. *et al.* Endogenous human microRNAs that suppress breast cancer metastasis. *Nature* **451**, 147–152 (2008).
374. Cheung, W. M., Fu, W. Y., Hui, W. S. & Ip, N. Y. Production of human CNS neurons from embryonal carcinoma cells using a cell aggregation method. *BioTechniques* **26**, 946–948, 950–952, 954 (1999).
375. Deb-Rinker, P., Ly, D., Jezierski, A., Sikorska, M. & Walker, P. R. Sequential DNA Methylation of the Nanog and Oct-4 Upstream Regions in Human NT2 Cells during Neuronal Differentiation. *Journal of Biological Chemistry* **280**, 6257 –6260 (2005).
376. Sikorska, M. *et al.* Epigenetic modifications of SOX2 enhancers, SRR1 and SRR2, correlate with in vitro neural differentiation. *Journal of Neuroscience Research* **86**, 1680–1693 (2008).
377. Morrisey, E. E., Ip, H. S., Lu, M. M. & Parmacek, M. S. GATA-6: a zinc finger transcription factor that is expressed in multiple cell lineages derived from lateral mesoderm. *Dev. Biol.* **177**, 309–322 (1996).
378. Greber, B., Lehrach, H. & Adjaye, J. Silencing of core transcription factors in human EC cells highlights the importance of autocrine FGF signaling for self-renewal. *BMC Dev. Biol* **7**, 46 (2007).
379. Matoba, R. *et al.* Dissecting Oct3/4-Regulated Gene Networks in Embryonic Stem Cells by Expression Profiling. *PLoS ONE* **1**, e26 (2006).
380. Morrisey, E. E. *et al.* GATA6 regulates HNF4 and is required for differentiation of visceral endoderm in the mouse embryo. *Genes Dev* **12**, 3579–3590 (1998).
381. Wang, Y., Smedberg, J. L., Cai, K. Q., Capo-Chichi, D. C. & Xu, X.-X. Ectopic expression of GATA6 bypasses requirement for Grb2 in primitive endoderm formation. *Dev. Dyn.* **240**, 566–576 (2011).
382. Futaki, S., Hayashi, Y., Emoto, T., Weber, C. N. & Sekiguchi, K. Sox7 Plays Crucial Roles in Parietal Endoderm Differentiation in F9 Embryonal Carcinoma Cells through Regulating Gata-4 and Gata-6 Expression. *Mol. Cell. Biol.* **24**, 10492–10503 (2004).
383. Kim, J. B. *et al.* Direct reprogramming of human neural stem cells by OCT4. *Nature* **461**, 649–643 (2009).
384. Ivey, K. N. *et al.* MicroRNA Regulation of Cell Lineages in Mouse and Human Embryonic Stem Cells. *Cell Stem Cell* **2**, 219–229 (2008).

385. Kee, K., Gonsalves, J. M., Clark, A. T. & Pera, R. A. R. Bone Morphogenetic Proteins Induce Germ Cell Differentiation from Human Embryonic Stem Cells. *Stem Cells and Development* **15**, 831–837 (2006).
386. Wilson, K. D. *et al.* MicroRNA Profiling of Human-Induced Pluripotent Stem Cells. *Stem Cells and Development* **18**, 749–757 (2009).
387. Panula, S. *et al.* Human germ cell differentiation from fetal- and adult-derived induced pluripotent stem cells. *Hum. Mol. Genet.* **20**, 752–762 (2011).
388. Gao, X. *et al.* A Role of N-Cadherin in Neuronal Differentiation of Embryonic Carcinoma P19 Cells. *Biochemical and Biophysical Research Communications* **284**, 1098–1103 (2001).
389. Chung, S., Arrell, D. K., Faustino, R. S., Terzic, A. & Dzeja, P. P. Glycolytic network restructuring integral to the energetics of embryonic stem cell cardiac differentiation. *Journal of Molecular and Cellular Cardiology* **48**, 725–734 (2010).
390. Elizalde, C. *et al.* Distinct Roles for Wnt-4 and Wnt-11 During Retinoic Acid-Induced Neuronal Differentiation. *STEM CELLS* **29**, 141–153 (2011).
391. Wagner-Ecker, M., Schwager, C., Wirkner, U., Abdollahi, A. & Huber, P. E. MicroRNA expression after ionizing radiation in human endothelial cells. *Radiation Oncology* **5**, 25 (2010).
392. Kano, M. *et al.* miR-145, miR-133a and miR-133b: Tumor-suppressive miRNAs target FSCN1 in esophageal squamous cell carcinoma. *International Journal of Cancer* **127**, 2804–2814 (2010).
393. Nohata, N. *et al.* miR-1 as a tumor suppressive microRNA targeting TAGLN2 in head and neck squamous cell carcinoma. *Oncotarget* **2**, 29–42
394. Ichimi, T. *et al.* Identification of novel microRNA targets based on microRNA signatures in bladder cancer. *International Journal of Cancer* **125**, 345–352 (2009).
395. Duester, G. Retinoic Acid Synthesis and Signaling during Early Organogenesis. *Cell* **134**, 921–931 (2008).
396. Wong, T.-S., Ho, W.-K., Chan, J. Y.-W., Ng, R. W.-M. & Wei, W. I. Mature miR-184 and squamous cell carcinoma of the tongue. *ScientificWorldJournal* **9**, 130–132 (2009).
397. Shu, M. *et al.* MiR-335 is Required for Differentiation of Malignant Glioma Cells Induced by Activation of cAMP/PKA Pathway. *Molecular Pharmacology* (2011).doi:10.1124/mol.111.076166
398. Wang, J. & Ruan, K. miR-335 is involved in the rat epididymal development by targeting the mRNA of RASA1. *Biochem. Biophys. Res. Commun* **402**, 222–227 (2010).
399. Talbot, S. G. *et al.* Gene Expression Profiling Allows Distinction between Primary and Metastatic Squamous Cell Carcinomas in the Lung. *Cancer Research* **65**, 3063 –3071 (2005).
400. Favreau, A. J. & Sathyanarayana, P. miR-590-5p, miR-219-5p, miR-15b and miR-628-5p are commonly regulated by IL-3, GM-CSF and G-CSF in acute myeloid leukemia. *Leukemia Research* (2011).doi:10.1016/j.leukres.2011.09.027
401. Zinn, P. O. *et al.* Radiogenomic Mapping of Edema/Cellular Invasion MRI-Phenotypes in Glioblastoma Multiforme. *PLoS ONE* **6**, e25451 (2011).
402. Rao, S. A. M., Santosh, V. & Somasundaram, K. Genome-wide expression profiling identifies deregulated miRNAs in malignant astrocytoma. *Mod. Pathol.* **23**, 1404–1417 (2010).
403. Saus, E. *et al.* Genetic variants and abnormal processing of pre-miR-182, a circadian clock modulator, in major depression patients with late insomnia. *Hum. Mol. Genet.* **19**, 4017–4025 (2010).
404. Kocerha, J. *et al.* MicroRNA-219 modulates NMDA receptor-mediated neurobehavioral dysfunction. *Proc. Natl. Acad. Sci. U.S.A* **106**, 3507–3512 (2009).
405. Gillis, A. J. M. *et al.* Expression and interdependencies of pluripotency factors LIN28, OCT3/4, NANOG and SOX2 in human testicular germ cells and tumours of the testis. *International Journal of Andrology* (2011).doi:10.1111/j.1365-2605.2011.01148.x
406. Scarola, M., Schoeftner, S., Schneider, C. & Benetti, R. miR-335 directly targets Rb1 (pRb/p105) in a proximal connection to p53-dependent stress response. *Cancer Res* **70**, 6925–6933 (2010).

407. Iliopoulos, D., Jaeger, S. A., Hirsch, H. A., Bulyk, M. L. & Struhl, K. STAT3 Activation of miR-21 and miR-181b-1 via PTEN and CYLD Are Part of the Epigenetic Switch Linking Inflammation to Cancer. *Molecular Cell* **39**, 493–506 (2010).
408. White, N. M. A. *et al.* miRNA profiling for clear cell renal cell carcinoma: biomarker discovery and identification of potential controls and consequences of miRNA dysregulation. *J. Urol.* **186**, 1077–1083 (2011).
409. Schmitz, K. J. *et al.* Differential expression of microRNA-675, microRNA-139-3p and microRNA-335 in benign and malignant adrenocortical tumours. *J. Clin. Pathol.* **64**, 529–535 (2011).
410. Soon, P. S. H. *et al.* miR-195 and miR-483-5p Identified as Predictors of Poor Prognosis in Adrenocortical Cancer. *Clin. Cancer Res.* **15**, 7684–7692 (2009).
411. Sorrentino, A. *et al.* Role of microRNAs in drug-resistant ovarian cancer cells. *Gynecologic oncology* **111**, 478–486 (2008).
412. Nakanishi, N. *et al.* The up-regulation of microRNA-335 is associated with lipid metabolism in liver and white adipose tissue of genetically obese mice. *Biochemical and Biophysical Research Communications* **385**, 492–496 (2009).
413. Kim, N. *et al.* Expression profiles of miRNAs in human embryonic stem cells during hepatocyte differentiation. *Hepatology Research* **41**, 170–183 (2011).
414. Ronchetti, D. *et al.* An integrative genomic approach reveals coordinated expression of intronic miR-335, miR-342, and miR-561 with deregulated host genes in multiple myeloma. *BMC Med Genomics* **1**, 37 (2008).
415. Sekkaï, D. *et al.* Microarray Analysis of LIF/Stat3 Transcriptional Targets in Embryonic Stem Cells. *STEM CELLS* **23**, 1634–1642 (2005).
416. Iliopoulos, D., Rotem, A. & Struhl, K. Inhibition of miR-193a Expression by Max and RXRα Activates K-Ras and PLAU to Mediate Distinct Aspects of Cellular Transformation. *Cancer Research* **71**, 5144 –5153 (2011).
417. Su, D. & Gudas, L. J. Retinoic acid receptor γ activates receptor tyrosine kinase Tie1 gene transcription through transcription factor GATA4 in F9 stem cells. *Experimental Hematology* **36**, 624–641 (2008).
418. Su, D. & Gudas, L. J. Gene expression profiling elucidates a specific role for RARγ in the retinoic acid-induced differentiation of F9 teratocarcinoma stem cells. *Biochemical Pharmacology* **75**, 1129–1160 (2008).
419. Boer, B. *et al.* Elevating the levels of Sox2 in embryonal carcinoma cells and embryonic stem cells inhibits the expression of Sox2:Oct-3/4 target genes. *Nucleic Acids Res.* **35**, 1773–1786 (2007).
420. Kopp, J. L., Ormsbee, B. D., Desler, M. & Rizzino, A. Small Increases in the Level of Sox2 Trigger the Differentiation of Mouse Embryonic Stem Cells. *STEM CELLS* **26**, 903–911 (2008).
421. Zhao, S., Nichols, J., Smith, A. G. & Li, M. SoxB transcription factors specify neuroectodermal lineage choice in ES cells. *Mol. Cell. Neurosci.* **27**, 332–342 (2004).
422. Ikushima, H. *et al.* Autocrine TGF-β Signaling Maintains Tumorigenicity of Glioma-Initiating Cells through Sry-Related HMG-Box Factors. *Cell Stem Cell* **5**, 504–514 (2009).
423. Mukhopadhyay, P. *et al.* MicroRNA gene expression signatures in the developing neural tube. *Birth Defects Research Part A: Clinical and Molecular Teratology* **91**, 744–762 (2011).
424. Williams, A. E., Moschos, S. A., Perry, M. M., Barnes, P. J. & Lindsay, M. A. Maternally imprinted microRNAs are differentially expressed during mouse and human lung development. *Developmental Dynamics* **236**, 572–580 (2007).
425. Hikasa, H. *et al.* Regulation of TCF3 by Wnt-dependent phosphorylation during vertebrate axis specification. *Dev. Cell* **19**, 521–532 (2010).
426. Merrill, B. J., Gat, U., DasGupta, R. & Fuchs, E. Tcf3 and Lef1 regulate lineage differentiation of multipotent stem cells in skin. *Genes Dev.* **15**, 1688–1705 (2001).
427. Kim, H. & Dorsky, R. I. Tcf7l1 is required for spinal cord progenitor maintenance. *Developmental Dynamics* **240**, 2256–2264 (2011).

428. Chen, Y., Li, D. & Li, S. The Alox5 gene is a novel therapeutic target in cancer stem cells of chronic myeloid leukemia. *Cell Cycle* **8**, 3488–3492 (2009).
429. Yilmaz, Ö. H. *et al.* Pten dependence distinguishes haematopoietic stem cells from leukaemia-initiating cells. *Nature* **441**, 475–482 (2006).
430. Guzman, M. L. *et al.* The sesquiterpene lactone parthenolide induces apoptosis of human acute myelogenous leukemia stem and progenitor cells. *Blood* **105**, 4163–4169 (2005).
431. Smith, K. M. *et al.* Selective targeting of neuroblastoma tumour-initiating cells by compounds identified in stem cell-based small molecule screens. *EMBO Molecular Medicine* **2**, 371–384 (2010).
432. Bomken, S., Fišer, K., Heidenreich, O. & Vormoor, J. Understanding the cancer stem cell. *Br J Cancer* **103**, 439–445 (2010).
433. Chambon, P. A decade of molecular biology of retinoic acid receptors. *FASEB J.* **10**, 940–954 (1996).
434. Blomhoff, R. & Blomhoff, H. K. Overview of retinoid metabolism and function. *Journal of Neurobiology* **66**, 606–630 (2006).
435. Jinek, M. & Doudna, J. A. A three-dimensional view of the molecular machinery of RNA interference. *Nature* **457**, 405–412 (2009).
436. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**, 806–811 (1998).
437. Rao, D. D., Vorhies, J. S., Senzer, N. & Nemunaitis, J. siRNA vs. shRNA: Similarities and differences. *Advanced Drug Delivery Reviews* **61**, 746–759 (2009).
438. The UniProt Consortium Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Research* **40**, D71–D75 (2011).
439. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
440. Liu, Y. *et al.* Genome wide profiling of human embryonic stem cells (hESCs), their derivatives and embryonal carcinoma cells to develop base profiles of U.S. Federal government approved hESC lines. *BMC Dev. Biol* **6**, 20 (2006).
441. Xie, Y. *et al.* Involvement of visinin-like protein-1 (VSNL-1) in regulating proliferative and invasive properties of neuroblastoma. *Carcinogenesis* **28**, 2122–2130 (2007).
442. Skotheim, R. I. *et al.* Differentiation of human embryonal carcinomas in vitro and in vivo reveals expression profiles relevant to normal development. *Cancer Res* **65**, 5588–5598 (2005).
443. Wickborn, C., Klein-Szanto, A. J., Schlag, P. M. & Braunewell, K.-H. Correlation of visinin-like-protein-1 expression with clinicopathological features in squamous cell carcinoma of the esophagus. *Mol. Carcinog* **45**, 572–581 (2006).
444. Synnergren, J. *et al.* Differentiating human embryonic stem cells express a unique housekeeping gene signature. *Stem Cells* **25**, 473–480 (2007).
445. Braunewell, K. H. & Gundelfinger, E. D. Low level expression of calcium-sensor protein VILIP induces cAMP-dependent differentiation in rat C6 glioma cells. *Neurosci. Lett* **234**, 139–142 (1997).
446. Gonzalez Guerrico, A. M. *et al.* Visinin-like protein-1 is a potent inhibitor of cell adhesion and migration in squamous carcinoma cells. *Oncogene* **24**, 2307–2316 (2005).
447. West, J. A. *et al.* A role for Lin28 in primordial germ-cell development and germ-cell malignancy. *Nature* **460**, 909–913 (2009).
448. Zuccotti, M. *et al.* Oct-4 regulates the expression of Stella and Foxj2 at the Nanog locus: implications for the developmental competence of mouse oocytes. *Hum. Reprod* **24**, 2225–2237 (2009).
449. Ezeh, U. I., Turek, P. J., Reijo, R. A. & Clark, A. T. Human embryonic stem cell genes OCT4, NANOG, STELLAR, and GDF3 are expressed in both seminoma and breast carcinoma. *Cancer* **104**, 2255–2265 (2005).

450. Huang, P. *et al.* Differentiation of human umbilical cord Wharton's jelly-derived mesenchymal stem cells into germ-like cells in vitro. *J. Cell. Biochem.* n/a–n/a (2009).doi:10.1002/jcb.22453
451. Tiedemann, R. E. *et al.* Kinome-wide RNAi studies in human multiple myeloma identify vulnerable kinase targets, including a lymphoid-restricted kinase, GRK6. *Blood* **115**, 1594–1604 (2010).
452. Chiu, T., Chang, J.-S., Jan, Y.-H., Lai, T. & Hsiao, M. Identification of AK3L1 as a New Lung Cancer Tumor Progression-associated Gene by Comparative Genomic and Proteomic Analysis. *FASEB J.* **24**, 354.3 (2010).
453. Jincho, Y. *et al.* Identification of Genes Aberrantly Expressed in Mouse Embryonic Stem Cell-Cloned Blastocysts. *Biology of Reproduction* **78**, 568–576 (2008).
454. Hedge, T. A. & Mason, I. Expression of Shisa2, a modulator of both Wnt and Fgf signaling, in the chick embryo. *Int. J. Dev. Biol* **52**, 81–85 (2008).
455. Zhu, Y. *et al.* Expression and roles of a xenopus head-forming gene homologue in human cancer cell lines. *Nagoya J Med Sci* **70**, 73–82 (2008).
456. Blick, T. *et al.* Epithelial Mesenchymal Transition Traits in Human Breast Cancer Cell Lines Parallel the CD44hi/CD24lo/- Stem Cell Phenotype in Human Breast Cancer. *J Mammary Gland Biol Neoplasia* **15**, 235–252 (2010).
457. Hjelmeland, A. B. *et al.* Acidic stress promotes a glioma stem cell phenotype. *Cell Death Differ* **18**, 829–840 (2011).
458. Loh, Y.-H. *et al.* The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet* **38**, 431–440 (2006).
459. Chew, J.-L. *et al.* Reciprocal Transcriptional Regulation of Pou5f1 and Sox2 via the Oct4/Sox2 Complex in Embryonic Stem Cells. *Molecular and Cellular Biology* **25**, 6031 –6046 (2005).
460. Masui, S. *et al.* Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat. Cell Biol* **9**, 625–635 (2007).
461. Graham, V., Khudyakov, J., Ellis, P. & Pevny, L. SOX2 Functions to Maintain Neural Progenitor Identity. *Neuron* **39**, 749–765 (2003).
462. Episkopou, V. SOX2 functions in adult neural stem cells. *Trends Neurosci.* **28**, 219–221 (2005).
463. Foshay, K. M. & Gallicano, G. I. Regulation of Sox2 by STAT3 initiates commitment to the neural precursor cell fate. *Stem Cells Dev.* **17**, 269–278 (2008).
464. Qu, Q. & Shi, Y. Neural Stem Cells in the Developing and Adult Brains. *J Cell Physiol* **221**, 5–9 (2009).
465. Szulwach, K. E. *et al.* Cross talk between microRNA and epigenetic regulation in adult neurogenesis. *J. Cell Biol.* **189**, 127–141 (2010).
466. Fang, X. *et al.* The SOX2 response program in glioblastoma multiforme: an integrated ChIP-seq, expression microarray, and microRNA analysis. *BMC Genomics* **12**, 11 (2011).
467. Looijenga, L. H. J. *et al.* POU5F1 (OCT3/4) Identifies Cells with Pluripotent Potential in Human Germ Cell Tumors. *Cancer Research* **63**, 2244 –2250 (2003).
468. de Jong, J. *et al.* Diagnostic value of OCT3/4 for pre-invasive and invasive testicular germ cell tumours. *The Journal of Pathology* **206**, 242–249 (2005).
469. Atlasi, Y., Mowla, S. J., Ziaee, S. A. M., Gokhale, P. J. & Andrews, P. W. OCT4 Spliced Variants Are Differentially Expressed in Human Pluripotent and Nonpluripotent Cells. *STEM CELLS* **26**, 3068–3074 (2008).
470. Jung, M. *et al.* A data integration approach to mapping OCT4 gene regulatory networks operative in embryonic stem cells and embryonal carcinoma cells. *PLoS ONE* **5**, e10709 (2010).
471. Liu, H. & Kohane, I. S. Tissue and Process Specific microRNA–mRNA Co-Expression in Mammalian Development and Malignancy. *PLoS ONE* **4**, e5436 (2009).
472. Nuovo, G. J. In situ detection of microRNAs in paraffin embedded, formalin fixed tissues and the co-localization of their putative targets. *Methods* **52**, 307–315 (2010).
473. Na, Y. J. *et al.* Comprehensive analysis of microRNA-mRNA co-expression in circadian rhythm. *Exp. Mol. Med.* **41**, 638–647 (2009).

474. Guimbellot, J. S. *et al.* Correlation of microRNA levels during hypoxia with predicted target mRNAs through genome-wide microarray analysis. *BMC Med Genomics* **2**, 15 (2009).
475. Nunez-Iglesias, J., Liu, C.-C., Morgan, T. E., Finch, C. E. & Zhou, X. J. Joint Genome-Wide Profiling of miRNA and mRNA Expression in Alzheimer's Disease Cortex Reveals Altered miRNA Regulation. *PLoS ONE* **5**, e8898 (2010).
476. Seo, J. & Hoffman, E. P. Probe set algorithms: is there a rational best bet? *BMC Bioinformatics* **7**, 395–395
477. Irizarry, R. A. *et al.* Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* **31**, e15 (2003).
478. Smyth, G. K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3**, Article3 (2004).
479. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44–57 (2009).
480. Chuaqui, R. F. *et al.* Post-analysis follow-up and validation of microarray experiments. *Nat. Genet.* **32 Suppl**, 509–514 (2002).
481. Tsuneyoshi, N. *et al.* PRDM14 suppresses expression of differentiation marker genes in human embryonic stem cells. *Biochem. Biophys. Res. Commun.* **367**, 899–905 (2008).
482. Yamaji, M. *et al.* Critical function of Prdm14 for the establishment of the germ cell lineage in mice. *Nature Genetics* **40**, 1016–1022 (2008).
483. Kim, S. *et al.* Identification of Developmental Pluripotency Associated 5 Expression in Human Pluripotent Stem Cells. *STEM CELLS* **23**, 458–462 (2005).
484. Barker, N. *et al.* Crypt stem cells as the cells-of-origin of intestinal cancer. *Nature* **457**, 608–611 (2008).
485. Vermeulen, L. *et al.* Single-cell cloning of colon cancer stem cells reveals a multi-lineage differentiation capacity. *PNAS* **105**, 13427–13432 (2008).
486. Chalamalasetty, R. B. *et al.* The Wnt3a/[beta]-catenin target gene Mesogenin1 controls the segmentation clock by activating a Notch signalling program. *Nature Communications* **2**, 390 (2011).
487. Tesar, P. J. *et al.* New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448**, 196–199 (2007).
488. Vodyanik, M. A. *et al.* A Mesoderm-Derived Precursor for Mesenchymal Stem and Endothelial Cells. *Cell Stem Cell* **7**, 718–729 (2010).
489. Katoh, M. & Katoh, M. CER1 is a common target of WNT and NODAL signaling pathways in human embryonic stem cells. *Int. J. Mol. Med.* **17**, 795–799 (2006).
490. Pereira, L. A. *et al.* Pdgfrα and Flk1 are direct target genes of Mixl1 in differentiating embryonic stem cells. *Stem Cell Res* **8**, 165–179 (2012).
491. Asiedu, M. K., Ingle, J. N., Behrens, M. D., Radisky, D. C. & Knutson, K. L. TGFβ/TNFα-Mediated Epithelial–Mesenchymal Transition Generates Breast Cancer Stem Cells with a Claudin-Low Phenotype. *Cancer Res* **71**, 4707–4719 (2011).
492. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. & Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **40**, D109–114 (2012).
493. Capecchi, M. R. Hox genes and mammalian development. *Cold Spring Harb. Symp. Quant. Biol.* **62**, 273–281 (1997).
494. Simeone, A. *et al.* Sequential activation of HOX2 homeobox genes by retinoic acid in human embryonal carcinoma cells. *, Published online: 23 August 1990; | doi:10.1038/346763a0* **346**, 763–766 (1990).
495. Simeone, A. *et al.* Differential regulation by retinoic acid of the homeobox genes of the four HOX loci in human embryonal carcinoma cells. *Mechanisms of Development* **33**, 215–227 (1991).
496. Mavilio, F., Simeone, A., Boncinelli, E. & Andrews, P. W. Activation of four homeobox gene clusters in human embryonal carcinoma cells induced to differentiate by retinoic acid. *Differentiation* **37**, 73–79 (1988).

497. Ueda, T. & Yoshida, M. HMGB proteins and transcriptional regulation. *Biochim. Biophys. Acta* **1799**, 114–118 (2010).
498. Palmer, R. D. *et al.* Malignant Germ Cell Tumors Display Common MicroRNA Profiles Resulting in Global Changes in Expression of Messenger RNA Targets. *Cancer Res* **70**, 2911–2923 (2010).
499. Cole, M. F., Johnstone, S. E., Newman, J. J., Kagey, M. H. & Young, R. A. Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. *Genes Dev* **22**, 746–755 (2008).
500. Maruyama, M., Ichisaka, T., Nakagawa, M. & Yamanaka, S. Differential Roles for Sox15 and Sox2 in Transcriptional Control in Mouse Embryonic Stem Cells. *J. Biol. Chem.* **280**, 24371–24379 (2005).
501. Murai, M. *et al.* Aberrant Methylation and Silencing of the BNIP3 Gene in Colorectal and Gastric Cancer. *Clin Cancer Res* **11**, 1021–1027 (2005).
502. Depaepe, V. *et al.* Ephrin signalling controls brain size by regulating apoptosis of neural progenitors. *Nature* **435**, 1244–1250 (2005).
503. Lahav, R. Endothelin receptor B is required for the expansion of melanocyte precursors and malignant melanoma. *Int. J. Dev. Biol.* **49**, 173–180 (2005).
504. Pollock, A. S., Turck, J. & Lovett, D. H. The prodomain of interleukin 1alpha interacts with elements of the RNA processing apparatus and induces apoptosis in malignant cells. *FASEB J.* **17**, 203–213 (2003).
505. Valencia, C. A., Cotten, S. W. & Liu, R. Cleavage of BNIP-2 and BNIP-XL by caspases. *Biochemical and Biophysical Research Communications* **364**, 495–501 (2007).
506. Kishi, S., Wulf, G., Nakamura, M. & Lu, K. P. Telomeric protein Pin2/TRF1 induces mitotic entry and apoptosis in cells with short telomeres and is down-regulated in human breast tumors. *Oncogene* **20**, 1497–1508 (2001).
507. Ju, J. *et al.* CD24 enhances DNA damage-induced apoptosis by modulating NF-κB signaling in CD44-expressing breast cancer cells. *Carcinogenesis* **32**, 1474–1483 (2011).
508. Eriksson, M. *et al.* Oncolytic Adenoviruses Kill Breast Cancer Initiating CD44+CD24-/Low Cells. *Mol Ther* **15**, 2088–2093 (2007).
509. Williams, T. A. *et al.* Teratocarcinoma-derived growth factor-1 is upregulated in aldosterone-producing adenomas and increases aldosterone secretion and inhibits apoptosis in vitro. *Hypertension* **55**, 1468–1475 (2010).
510. He, W. *et al.* TLR4 signaling promotes immune escape of human lung cancer cells by inducing immunosuppressive cytokines and apoptosis resistance. *Mol. Immunol.* **44**, 2850–2859 (2007).
511. Asano, J. *et al.* The serine/threonine kinase Pim-2 is a novel anti-apoptotic mediator in myeloma cells. *Leukemia* **25**, 1182–1188 (2011).
512. Alvero, A. B. *et al.* Molecular phenotyping of human ovarian cancer stem cells unravels the mechanisms for repair and chemoresistance. *Cell Cycle* **8**, 158–166 (2009).
513. Kim, K., Lu, Z. & Hay, E. D. DIRECT EVIDENCE FOR A ROLE OF β-CATENIN/LEF-1 SIGNALING PATHWAY IN INDUCTION OF EMT. *Cell Biology International* **26**, 463–476 (2002).
514. Kang, M. H., Kim, J. S., Seo, J. E., Oh, S. C. & Yoo, Y. A. BMP2 accelerates the motility and invasiveness of gastric cancer cells via activation of the phosphatidylinositol 3-kinase (PI3K)/Akt pathway. *Exp. Cell Res.* **316**, 24–37 (2010).
515. Arnold, S. J., Hofmann, U. K., Bikoff, E. K. & Robertson, E. J. Pivotal roles for eomesodermin during axis formation, epithelium-to-mesenchyme transition and endoderm specification in the mouse. *Development* **135**, 501–511 (2008).
516. Hajra, K. M., Chen, D. Y.-S. & Fearon, E. R. The SLUG Zinc-Finger Protein Represses E-Cadherin in Breast Cancer. *Cancer Research* **62**, 1613 –1618 (2002).
517. Wong, Y.-H. *et al.* Protogenin Defines a Transition Stage during Embryonic Neurogenesis and Prevents Precocious Neuronal Differentiation. *The Journal of Neuroscience* **30**, 4428 –4439 (2010).

518. Vergaño-Vera, E., Fernández, A. P., Hurtado-Chong, A., Vicario-Abejón, C. & Martínez, A. Lack of adrenomedullin affects growth and differentiation of adult neural stem/progenitor cells. *Cell and Tissue Research* **340**, 1–11 (2010).
519. Bloom, L. & Horvitz, H. R. The Caenorhabditis elegans gene unc-76 and its human homologs define a new gene family involved in axonal outgrowth and fasciculation. *PNAS* **94**, 3414–3419 (1997).
520. Gualandris, A. *et al.* Microenvironment drives the endothelial or neural fate of differentiating embryonic stem cells coexpressing neuropilin-1 and Flk-1. *FASEB J.* **23**, 68–78 (2009).
521. Ferlin, A. *et al.* Variants in KITLG predispose to testicular germ cell cancer independently from spermatogenic function. *Endocr. Relat. Cancer* **19**, 101–108 (2012).
522. Califano, D. *et al.* FEZ1/LZTS1 protein expression in ovarian cancer. *Journal of Cellular Physiology* **222**, 382–386 (2010).
523. Marson, A. *et al.* Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* **134**, 521–533 (2008).
524. Mukherjee, N. *et al.* Integrative Regulatory Mapping Indicates that the RNA-Binding Protein HuR Couples Pre-mRNA Processing and mRNA Stability. *Molecular Cell* **43**, 327–339 (2011).
525. Xin, F. *et al.* Computational analysis of microRNA profiles and their target genes suggests significant involvement in breast cancer antiestrogen resistance. *Bioinformatics* **25**, 430–434 (2009).
526. Nam, S., Kim, B., Shin, S. & Lee, S. miRGator: an integrated system for functional annotation of microRNAs. *Nucleic Acids Research* **36**, D159–D164 (2007).
527. Wang, X. miRDB: A microRNA target prediction and functional annotation database with a wiki interface. *RNA* **14**, 1012–1017 (2008).
528. Guo, H., Ingolia, N. T., Weissman, J. S. & Bartel, D. P. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466**, 835–840 (2010).
529. Baek, D. *et al.* The impact of microRNAs on protein output. *Nature* **455**, 64–71 (2008).
530. Shahab, S. W. *et al.* Evidence for the Complexity of MicroRNA-Mediated Regulation in Ovarian Cancer: A Systems Approach. *PLoS ONE* **6**, e22508 (2011).
531. Ma, L. *et al.* An Integrated Analysis of miRNA and mRNA Expressions in Non-Small Cell Lung Cancers. *PLoS One* **6**, (2011).
532. Su, W.-L., Kleinhanz, R. R. & Schadt, E. E. Characterizing the role of miRNAs within gene regulatory networks using integrative genomics techniques. *Mol Syst Biol* **7**, 490 (2011).
533. Ambs, S. *et al.* Genomic Profiling of MicroRNA and Messenger RNA Reveals Deregulated MicroRNA Expression in Prostate Cancer. *Cancer Research* **68**, 6162 –6170 (2008).
534. Moser, J. J. & Fritzler, M. J. The MicroRNA and MessengerRNA Profile of the RNA-Induced Silencing Complex in Human Primary Astrocyte and Astrocytoma Cells. *PLoS One* **5**, (2010).
535. Selbach, M. *et al.* Widespread changes in protein synthesis induced by microRNAs. *Nature* **455**, 58–63 (2008).
536. Houldsworth, J., Heath, S. C., Bosl, G. J., Studer, L. & Chaganti, R. S. K. Expression profiling of lineage differentiation in pluripotential human embryonal carcinoma cells. *Cell Growth Differ* **13**, 257–264 (2002).
537. Zhang, J. *et al.* Identification of an ES cell pluripotent state-specific DUSP6 enhancer. *Biochem. Biophys. Res. Commun.* **378**, 319–323 (2009).
538. Xiao, Q. *et al.* Nrf3-Pla2g7 Interaction Plays an Essential Role in Smooth Muscle Cell Differentiation From Stem Cells. *Arteriosclerosis, Thrombosis, and Vascular Biology* (2012).doi:10.1161/ATVBAHA.111.243188
539. Chen, B.-Z. *et al.* Identification of microRNAs expressed highly in pancreatic islet-like cell clusters differentiated from human embryonic stem cells. *Cell Biol. Int.* **35**, 29–37 (2011).
540. Zampetaki, A., Xiao, Q., Zeng, L., Hu, Y. & Xu, Q. TLR4 expression in mouse embryonic stem cells and in stem cell-derived vascular cells is regulated by epigenetic modifications. *Biochem. Biophys. Res. Commun.* **347**, 89–99 (2006).
541. Urbach, A. & Benvenisty, N. Studying Early Lethality of 45,XO (Turner's Syndrome) Embryos Using Human Embryonic Stem Cells. *PLoS ONE* **4**, e4175 (2009).

542. Gu, P. *et al.* Orphan nuclear receptor LRH-1 is required to maintain Oct4 expression at the epiblast stage of embryonic development. *Mol. Cell. Biol.* **25**, 3492–3505 (2005).
543. Skottman, H. *et al.* Gene Expression Signatures of Seven Individual Human Embryonic Stem Cell Lines. *STEM CELLS* **23**, 1343–1356 (2005).
544. Wang, P., Rodriguez, R. T., Wang, J., Ghodasara, A. & Kim, S. K. Targeting SOX17 in Human Embryonic Stem Cells Creates Unique Strategies for Isolating and Analyzing Developing Endoderm. *Cell Stem Cell* **8**, 335–346 (2011).
545. Grabel, L., Becker, S., Lock, L., Maye, P. & Zanders, T. Using EC and ES cell culture to study early development: recent observations on Indian hedgehog and Bmps. *Int. J. Dev. Biol.* **42**, 917–925 (1998).
546. Cao, L. *et al.* Intestinal lineage commitment of embryonic stem cells. *Differentiation* **81**, 1–10 (2011).
547. Moon, S.-H. *et al.* Gene expression profiles in CHA3 and CHA4 human embryonic stem cells and embryoid bodies. *Mol. Cells* **31**, 315–326 (2011).
548. O'Neill, L. P., VerMilyea, M. D. & Turner, B. M. Epigenetic characterization of the early embryo with a chromatin immunoprecipitation protocol applicable to small cell populations. *Nat Genet* **38**, 835–841 (2006).
549. Wong, J. C. . *et al.* Definitive endoderm derived from human embryonic stem cells highly express the integrin receptors αV and β5. *Cell Adh Migr* **4**, 39–45 (2010).
550. Laslett, A. L. *et al.* Transcriptional analysis of early lineage commitment in human embryonic stem cells. *BMC Dev. Biol.* **7**, 12 (2007).
551. Ledran, M. H. *et al.* Efficient Hematopoietic Differentiation of Human Embryonic Stem Cells on Stromal Cells Derived from Hematopoietic Niches. *Cell Stem Cell* **3**, 85–98 (2008).
552. Marchand, M. *et al.* Transcriptomic signature of trophoblast differentiation in a human embryonic stem cell model. *Biol. Reprod.* **84**, 1258–1271 (2011).
553. Xiao, L., Yuan, X. & Sharkis, S. J. Activin A Maintains Self-Renewal and Regulates Fibroblast Growth Factor, Wnt, and Bone Morphogenic Protein Pathways in Human Embryonic Stem Cells. *STEM CELLS* **24**, 1476–1486 (2006).
554. D'Amour, K. A. *et al.* Efficient differentiation of human embryonic stem cells to definitive endoderm. *Nat Biotech* **23**, 1534–1541 (2005).
555. Shiraki, N. *et al.* Differentiation and characterization of embryonic stem cells into three germ layers. *Biochemical and Biophysical Research Communications* **381**, 694–699 (2009).
556. Aghajanova, L. *et al.* Expression of leukemia inhibitory factor and its receptors is increased during differentiation of human embryonic stem cells. *Fertility and Sterility* **86**, 1193–1209 (2006).
557. Schnetz, M. P. *et al.* CHD7 targets active gene enhancer elements to modulate ES cell-specific gene expression. *PLoS Genet.* **6**, e1001023 (2010).
558. Skottman, H. *et al.* Unique Gene Expression Signature by Human Embryonic Stem Cells Cultured Under Serum-Free Conditions Correlates with Their Enhanced and Prolonged Growth in an Undifferentiated Stage. *Stem Cells* **24**, 151–167 (2006).
559. Baker, R. K., Vanderboom, A. K., Bell, G. W. & Antin, P. B. Expression of the receptor tyrosine kinase gene EphB3 during early stages of chick embryo development. *Mechanisms of Development* **104**, 129–132 (2001).
560. Young, A. *et al.* Ion channels and ionotropic receptors in human embryonic stem cell derived neural progenitors. *Neuroscience* **192**, 793–805 (2011).
561. Redmer, T. *et al.* E-cadherin is crucial for embryonic stem cell pluripotency and can replace OCT4 during somatic cell reprogramming. *EMBO Rep* **12**, 720–726 (2011).
562. Evans-Anderson, H. J., Alfieri, C. M. & Yutzey, K. E. Regulation of Cardiomyocyte Proliferation and Myocardial Growth During Development by FOXO Transcription Factors. *Circulation Research* **102**, 686 –694 (2008).
563. Fardaei, M. *et al.* Three proteins, MBNL, MBLL and MBXL, co-localize in vivo with nuclear foci of expanded-repeat transcripts in DM1 and DM2 cells. *Hum. Mol. Genet.* **11**, 805–814 (2002).

564. Huang, H., Wahlin, K. J., McNally, M., Irving, N. D. & Adler, R. Developmental regulation of muscleblind-like (MBNL) gene expression in the chicken embryo retina. *Developmental Dynamics* **237**, 286–296 (2008).
565. Kang, J.-S., Mulieri, P. J., Hu, Y., Taliana, L. & Krauss, R. S. BOC, an Ig superfamily member, associates with CDO to positively regulate myogenic differentiation. *EMBO J.* **21**, 114–124 (2002).
566. Schlange, T., Andrée, B., Arnold, H. & Brand, T. Expression analysis of the chicken homologue of CITED2 during early stages of embryonic development. *Mech. Dev.* **98**, 157–160 (2000).
567. Wilczynska, K. M. *et al.* Nuclear Factor I Isoforms Regulate Gene Expression During the Differentiation of Human Neural Progenitors to Astrocytes. *STEM CELLS* **27**, 1173–1181 (2009).
568. Yu, Z. *et al.* The Grainyhead-like epithelial transactivator Get-1/Grhl3 regulates epidermal terminal differentiation and interacts functionally with LMO4. *Dev. Biol.* **299**, 122–136 (2006).
569. Silva, J., Wang, G. & Cowell, J. K. The temporal and spatial expression pattern of the LGI1 epilepsy predisposition gene during mouse embryonic cranial development. *BMC Neurosci* **12**, 43 (2011).
570. Placantonakis, D. G. *et al.* BAC Transgenesis in Human Embryonic Stem Cells as a Novel Tool to Define the Human Neural Lineage. *STEM CELLS* **27**, 521–532 (2009).
571. Lin, L. & Isacson, O. Axonal Growth Regulation of Fetal and Embryonic Stem Cell-Derived Dopaminergic Neurons by Netrin-1 and Slits. *STEM CELLS* **24**, 2504–2513 (2006).
572. Khokha, M. K., Hsu, D., Brunet, L. J., Dionne, M. S. & Harland, R. M. Gremlin is the BMP antagonist required for maintenance of Shh and Fgf signals during limb patterning. *Nature Genetics* **34**, 303–307 (2003).
573. Moore, D. J., Chambers, J. K., Murdock, P. R. & Emson, P. C. Human Ntera-2/D1 neuronal progenitor cells endogenously express a functional P2Y1 receptor. *Neuropharmacology* **43**, 966–978 (2002).
574. Megiorni, F., Mora, B., Indovina, P. & Mazzilli, M. C. Expression of neuronal markers during NTera2/cloneD1 differentiation by cell aggregation method. *Neuroscience Letters* **373**, 105–109 (2005).
575. Good, P. *et al.* The HumanMusashi Homolog 1(MSI1) Gene Encoding the Homologue of Musashi/Nrp-1, a Neural RNA-Binding Protein Putatively Expressed in CNS Stem Cells and Neural Progenitor Cells. *Genomics* **52**, 382–384 (1998).
576. Knecht, A. k., Good, P. j., Dawid, I. b. & Harland, R. m. Dorsal-ventral patterning and differentiation of noggin-induced neural tissue in the absence of mesoderm. *Development* **121**, 1927 –1935 (1995).
577. Mustain, W. C., Rychahou, P. G. & Evers, B. M. The role of neurotensin in physiologic and pathologic processes. *Curr Opin Endocrinol Diabetes Obes* **18**, 75–82 (2011).
578. Kele, J. *et al.* SFRP1 and 2 Dose-Dependently Regulate Midbrain Dopamine Neuron Development In vivo and in Embryonic Stem Cells. *Stem Cells (Dayton, Ohio)* (2012).doi:10.1002/stem.1049
579. Cajánek, L. *et al.* Wnt/beta-catenin signaling blockade promotes neuronal induction and dopaminergic differentiation in embryonic stem cells. *Stem Cells* **27**, 2917–2927 (2009).
580. Zhou, A.-D. *et al.* β-Catenin/LEF1 transactivates the microRNA-371-373 cluster that modulates the Wnt/β-catenin-signaling pathway. *Oncogene* (2011).doi:10.1038/onc.2011.461
581. Li, C., Scott, D. A., Hatch, E., Tian, X. & Mansour, S. L. Dusp6 (Mkp3) is a negative feedback regulator of FGF-stimulated ERK signaling during mouse development. *Development* **134**, 167 –176 (2007).
582. Bin Liu Association of the Dusp6 (Mkp3) Gene With Mouse Brain Weight and Forebrain Structure. *Journal of Child Neurology* **23**, 624 –627 (2008).
583. Okuda, Y., Ogura, E., Kondoh, H. & Kamachi, Y. B1 SOX Coordinate Cell Specification with Patterning and Morphogenesis in the Early Zebrafish Embryo. *PLoS Genet* **6**, (2010).

584. Rodda, D. J. *et al.* Transcriptional Regulation of Nanog by OCT4 and SOX2. *Journal of Biological Chemistry* **280**, 24731 –24737 (2005).
585. Matsumata, M., Uchikawa, M., Kamachi, Y. & Kondoh, H. Multiple N-cadherin enhancers identified by systematic functional screening indicate its Group B1 SOX-dependent regulation in neural and placodal development. *Developmental Biology* **286**, 601–617 (2005).
586. Engelen, E. *et al.* Sox2 cooperates with Chd7 to regulate genes that are mutated in human syndromes. *Nat Genet* **43**, 607–611 (2011).
587. Di Rocco, G. *et al.* The Recruitment of SOX/OCT Complexes and the Differential Activity of HOXA1 and HOXB1 Modulate the Hoxb1Auto-regulatory Enhancer Function. *Journal of Biological Chemistry* **276**, 20506 –20515 (2001).
588. Park, S. B. *et al.* SOX2 has a crucial role in the lineage determination and proliferation of mesenchymal stem cells through Dickkopf-1 and c-MYC. *Cell Death Differ.* (2011).doi:10.1038/cdd.2011.137
589. Dettman, E. J. & Justice, M. J. The Zinc Finger SET Domain Gene Prdm14 Is Overexpressed in Lymphoblastic Lymphomas with Retroviral Insertions at Evi32. *PLoS ONE* **3**, e3823 (2008).
590. Sharov, A. A. *et al.* Identification of Pou5f1, Sox2, and Nanog downstream target genes with statistical confidence by applying a novel algorithm to time course microarray and genome-wide chromatin immunoprecipitation data. *BMC Genomics* **9**, 269 (2008).
591. Stadler, B. *et al.* Characterization of microRNAs involved in embryonic stem cell states. *Stem Cells Dev.* **19**, 935–950 (2010).
592. Swarbrick, A. *et al.* miR-380-5p represses p53 to control cellular survival and is associated with poor outcome in MYCN amplified neuroblastoma. *Nat Med* **16**, 1134–1140 (2010).
593. Sun, J. *et al.* Microarray-based analysis of microRNA expression in breast cancer stem cells. *J Exp Clin Cancer Res* **29**, 174 (2010).
594. Yu, F. *et al.* Mir-30 reduction maintains self-renewal and inhibits apoptosis in breast tumor-initiating cells. *Oncogene* **29**, 4194–4204 (2010).
595. Godlewski, J., Newton, H. B., Chiocca, E. A. & Lawler, S. E. MicroRNAs and glioblastoma; the stem cell connection. *Cell Death Differ* **17**, 221–228 (2009).
596. Ciafrè, S. A. *et al.* Extensive modulation of a set of microRNAs in primary glioblastoma. *Biochemical and Biophysical Research Communications* **334**, 1351–1358 (2005).
597. Voorhoeve, P. M. *et al.* A Genetic Screen Implicates miRNA-372 and miRNA-373 As Oncogenes in Testicular Germ Cell Tumors. *Cell* **124**, 1169–1181 (2006).
598. Leis, O. *et al.* Sox2 expression in breast tumours and activation in breast cancer stem cells. *Oncogene* (2011).doi:10.1038/onc.2011.338
599. Xiang, R. *et al.* Downregulation of transcription factor SOX2 in cancer stem cells suppresses growth and metastasis of lung cancer. *Br. J. Cancer* **104**, 1410–1417 (2011).
600. Watanabe, K. *et al.* Cripto-1 Is a Cell Surface Marker for a Tumorigenic, Undifferentiated Subpopulation in Human Embryonal Carcinoma Cells. *STEM CELLS* **28**, 1303–1314 (2010).
601. Gallagher, M. *et al.* Regulation of microRNA biosynthesis and expression in 2102Ep embryonal carcinoma stem cells is mirrored in ovarian serous adenocarcinoma patients. *Journal of Ovarian Research*
602. Popovic, R. *et al.* Regulation of mir-196b by MLL and its overexpression by MLL fusions contributes to immortalization. *Blood* **113**, 3314–3322 (2009).
603. Lakomy, R. *et al.* MiR-195, miR-196b, miR-181c, miR-21 expression levels and O-6-methylguanine-DNA methyltransferase methylation status are associated with clinical outcome in glioblastoma patients. *Cancer Science* **102**, 2186–2190 (2011).
604. Bhatia, S., Kaul, D. & Varma, N. Functional genomics of tumor suppressor miR-196b in T-cell acute lymphoblastic leukemia. *Mol. Cell. Biochem.* **346**, 103–116 (2011).
605. Bhatia, S., Kaul, D. & Varma, N. Potential tumor suppressive function of miR-196b in B-cell lineage acute lymphoblastic leukemia. *Mol. Cell. Biochem.* **340**, 97–106 (2010).
606. Li, Y. *et al.* Ratio of miR-196s to HOXC8 messenger RNA correlates with breast cancer cell migration and metastasis. *Cancer Res.* **70**, 7894–7904 (2010).