**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.

# Cancer Stemness: Unravelling the Molecular Mechanisms Controlling Stem Cell Differentiation and Self-Renewal


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

**2009**

**By**

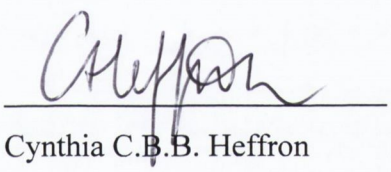
**Cynthia Constance Bridgette Beatrice Heffron**
**MB BCh BAO, BMedSci, PG Dip Stats**

**Supervisor:**
**Professor John J. O'Leary**
**MD, DPhil, MSc, FRCPath**

## Declaration

The work within this thesis is being submitted for the degree of PhD. I declare that this thesis has not been submitted as an exercise for a degree at this or any other University prior to now. I declare that it is the original work of this author. I hereby agree to allow the Library to lend or copy the thesis upon request.

This work was supervised by Professor John J. O'Leary, Department of Histopathology and Morbid Anatomy at Trinity College Dublin.



Cynthia C.B.B. Heffron

## Dedication

I would like to dedicate this thesis to my father whose love of research and delving into the unknown inspired me to undertake this body of work.

## Summary

Understanding the molecular mechanisms behind cancer cell proliferation and regulation is the key to discovering potential future successful therapies for cancer. The World Health Organisation estimates that 7.6 million people died from cancer in 2005 with that number expected to reach 9 million in 2015 and 11.5 million in 2030. The concept of a cancer stem cell as the core cancer initiating agent within a tumour has come to the forefront of tumour biology with a cancer stem cell population identified in many different tumour types including prostate, breast, skin, colon and ovarian cancer. Cancer stem cells and normal stem cells share two fundamental properties, namely the ability to self-renew and the ability to differentiate. The characterisation of the molecular mechanisms underlying these two processes will be vital to the understanding of the regulation of cancer cell proliferation and progression. The specific targeting of cancer stem cells within a tumour as a treatment modality may provide more effective cancer treatment in the future as many current treatments are not eliminating the small but significant population of cancer stem cells that result in metastasis and recurrence. Teratocarcinomas are malignant germ cell tumours that occur in the testes and ovary and are considered to be the classical stem cell tumour being composed of an undifferentiated population of embryonal carcinoma cells and mature differentiated tissues. This tumour's ability to differentiate as well as being inherently malignant makes teratoma tumourigenesis an ideal model to study the concept of cancer stemness and thus identify stem cell markers and targets that may be useful in the assessment and treatment of cancer stem cells in many other tumours.

This study sets out to investigate the similarities and differences between mouse embryonic stem cells and their malignant equivalent, embryonal carcinoma cells in an attempt to outline the fundamental molecular properties of self-renewal and differentiation within stem cells and cancer stem cells. Obtaining a transcriptome profile of the cancer stem cell or cancer stemness should provide new insight into the molecular mechanisms governing the cancer stem cell as well as highlighting new targets for future cancer therapies and for further study.

A mouse cell culture model of teratoma tumourigenesis was established using mouse embryonic stem cells, mouse pluripotent and nullipotent teratocarcinoma cells. Samples were harvested from each cell line in the undifferentiated state at day 0 and day 3 together with samples from day 3 post spontaneous differentiation. A similar human model was also constructed using human pluri- and nullipotent teratocarcinoma cells without the use of human embryonic stem cells, and therefore data was compared with that published in the literature. Gene expression profiles were generated using whole genome cDNA array analysis from Applied Biosystems and subsequent statistical analysis. A selection of gene targets was validated by quantitative real-time PCR.

Data obtained in this thesis from the comparison of benign and malignant stem cells has confirmed some of the current findings in the literature as well as highlighting many novel genes and their related pathways in the cancer stem cell. In particular, the importance of the extracellular matrix and stem cell niche has been reinforced by the data with a number of novel genes found to be differentially regulated in stem cells that could potentially revolutionise future cancer management strategies. Oxidative stress was revealed as an extremely significant pathway in stem cells with a unique group of genes encoding for dual specificity phosphatases (DUSPs) being significantly differentially regulated in cancer stem cells. Further transcriptome profiles of differentiation in the benign group (mES cells) and the malignant group (pluri- and nullipotent EC cells) were obtained allowing the determination of genes unique to the normal stem cell and the cancer stem cell. As a significant risk in the successful elimination of cancer stem cells is the coincidental eradication of normal stem cells, determining properties exclusive to the cancer stem cell is of vital importance to future cancer treatments.

This thesis offers a unique transcriptome profile of the cancer stem cell as well as a unique profile of those genes solely involved in cancer stem cell differentiation using both a mouse and human model of teratoma tumourigenesis. Multiple novel gene targets and signalling pathways have been identified meriting further study and analysis.

# Table of Contents

# List of Figures

# List of Tables

# Acknowledgements

I would like to thank my supervisor, Professor John O'Leary, for all his help and guidance throughout the course of my research work. His support and guidance were invaluable and without him this thesis would not be here in print.

I would also like to thank all in the Molecular Pathology Research Laboratory at the Coombe Women's Hospital, Dublin 8, in particular Dr. Michael Gallagher and Dr. Cara Martin for all their technical expertise and skilled guidance throughout my time at the Coombe and subsequently during the writing process. I would like to thank Amanda for all her little tips, advice and words of encouragement during the writing up process, they were extremely helpful. I would like to thank Ciara for being so supportive throughout.

I would like to thank my family for their unerring support throughout my career, research and indeed in everything I do and in particular my father, to whom this thesis is dedicated. He has been my inspiration in undertaking this work and was very positive and encouraging when I needed it most. Last and by no means least, I would like to thank my ever supportive husband, Ciarán, who was always there with uplifting words, support and with masses of confidence in my ability to see this work though to the end.

This work was part funded by Applied Biosystems.

# List of Abbreviations

| | |
|---|---|
| AB | Applied Biosystems |
| ABarray | Applied Biosystems Genome Survey Array Data Analysis program |
| ABC | Activated B cell |
| ABL | Stands for 'Abelson', the leukaemia virus its named after |
| ADAMTS | Adamalysin with thrombospondin type 1 motifs |
| AFP | Alpha-foeto-protein |
| AKAP | A-kinase anchor protein |
| AML | Acute myeloid leukaemia |
| ANOVA | Analysis of variance |
| ATC | Anaplastic thyroid cancer |
| ATCC | American Type Culture Collection (www.atcc.org) |
| β2-AR | β2 – adrenergic receptor |
| BCC | Basal cell carcinoma |
| BCR | Breakpoint cluster region |
| β-ME | β-mercaptoethanol |
| BMPs | Bone morphogenetic proteins |
| CA | California |
| CCD | Charge coupled device |
| CD | Cluster designation |
| cDNA | Complementary DNA |
| cfu | Colony-forming units |
| CML | Chronic myeloid leukaemia |
| CNS | Central nervous system |
| Conc. | Concentration |
| CSC | Cancer stem cells |
| CsCl | Cesium chloride |
| $C_T$ | Cycle threshold |
| CV | Covariance |
| DIG | Digoxigenin |
| DIG-11-UTP | Digoxigenin-11-uridine-5'-triphosphate |

| | |
|---|---|
| DKKs | Dickkoptfs |
| DLBCL | Diffuse large B cell lymphoma |
| DMEM | Dulbecco's modified Eagles Medium |
| DNA | Deoxy Nucleic Acid |
| dNTPs | Deoxyribonucleotide triphospates |
| DSCAM | Down syndrome cell adhesion molecule |
| DSMO | Dimethyl sulphoxide |
| DUSPs | Dual specificity phosphatases |
| dUTP | 2'-Deoxyuridine 5'-Triphosphate |
| EB | Embryoid bodies |
| EC cells | Embryonal carcinoma cells |
| ECM | Extracellular matrix |
| EDTA | Ethylenediaminetetraacetic acid |
| ELISA | Enzyme-linked immunosorbant assay |
| ES cells | Embryonic stem cells |
| EST | Expressed sequence tags |
| FAM | 6-carboxyfluorescine |
| FBS | Foetal bovine serum |
| Fbln | Fibulin |
| FC | Fold change |
| FDA | US Food and Drug Administration |
| FDR | False discovery rate |
| FGFs | Fibroblast growth factors |
| GAPDH | Glyceraldehyde 3-phosphate dehydrogenase |
| GCB | Germinal centre B-cell like |
| GP | Glycoprotein |
| hEC | Human embryonal carcinoma |
| ECM | Extracellular matrix |
| hESCs | Human embryonic stem cells |
| hES cells | Human embryonic stem cells |
| HgPRT | Hypoxanthine-guanine phosphoribosyl transferase |
| HMBA | Hexamethylene bisacetamide |
| HMG | High mobility group |

| | |
|---|---|
| HOX | Homeobox |
| HPGD | Hydroxyprostaglandin dehydrogenase |
| Hr | Hour |
| ICP | Internal control probes |
| ICT | Internal control target |
| ICM | Inner cell mass |
| ID | Identification |
| IGF | Insulin-like growth factor |
| IGFBP | Insulin-like growth factor binding protein |
| IL | Interleukin |
| Inc | Incorporated |
| IU | International units |
| IVT | In vitro transcription |
| JAK | Tyrosine kinase Janus |
| LiCl | Lithium chloride |
| LIF | Leukaemia inhibitory factor |
| LRP | Low density lipoprotein receptor-related protein |
| LSC | Leukaemic stem cells |
| Ltd | Limited |
| MA plot | M is the intensity ratio and A is the average intensity for a dot in the plot |
| MAP | Mitogen activated protein |
| MAQC | MicroArray Quality Control |
| MEF | Mitotically inactive feeders |
| mEC | Mouse embryonal carcinoma |
| mES cells | Mouse embryonic stem cells |
| mESCs | Mouse embryonic stem cells |
| MGB | Minor groove binders |
| miRNA | Micro RNAs |
| MIAME | Minimum Information About a Microarray Experiment |
| MKPs | MAP kinase phosphatases |
| MPSS | Massively parallel signature sequencing |
| mRNA | Messenger RNA |
| MuLVRT | Murine leukaemia virus reverse transcriptase |

| | |
|---|---|
| NCAM | Neural cell adhesion molecule 1 |
| NFQ | Non fluorescent quencher |
| NK | Natural killer |
| NSGT | Non seminomatous germ cell tumours |
| NTC | No template control |
| OAS | 2'-5'-oligoadenylate synthetase |
| OASL | OAS like |
| PANTHER | **P**rotein **AN**alysis **TH**rough **E**volutionary **R**elationships (www.pantherdb.org) |
| PBS | Phosphate buffered saline |
| PCR | Polymerase chain reaction |
| PGCs | Primordial germ cells |
| PI3K | Phosphoinositide 3-kinase |
| PKA | Protein kinase A |
| Prame | Preferentially expressed antigen in melanoma |
| Pramel | Preferentially expressed antigen in melanoma like |
| PTEN | Phosphatase and tensin homologue deleted from chromosome 10 |
| QA | Quality assurance |
| QC | Quality control |
| qRT-PCR | Quantitative reverse transcription polymerase chain reaction |
| RA | Retinoic acid |
| RIN | RNA integrity number |
| RNA | Ribo Nucleic Acid |
| RNAi | RNA interference |
| Rpm | Revolutions per minute |
| RT | Reverse transcription |
| RT-PCR | Reverse transcription polymerase chain reaction |
| SAGE | Serial analysis of gene expression |
| SFRPs | Frizzled-related proteins |
| SHH | Sonic hedgehog |
| SiRNA | Short interfering RNAs |
| SMAD | Similar to mothers against decapentaplegic homologue |
| S/N | Signal to noise |
| Sox | SRY-related HMG-box |

| | |
|---|---|
| SRY | Sex determining region Y |
| SSEA | Stage-specific embryonic antigen |
| STAT | Signal transducer and activator of transcription |
| TBE | Tris Borate EDTA |
| TGF | Transforming growth factor |
| TIK | Antiphosphotyrosine immmunoreactive kinase |
| TMB | Tetramethylbenzidine |
| TMC | Transmembrane channel |
| TNFR | Tumour necrosis factor receptor |
| TNFRSF | Tumour necrosis factor receptor superfamily |
| TQ | TaqMan® |
| UK | United Kingdom |
| UNG | Uracil-DNA glycosylase |
| USA | United States of America |
| UTP | Uridine 5'-Triphosphate |
| UV | Ultraviolet |
| VEGFR2 | Vascular endothelial growth factor receptor 2 |
| WHO | World Health Organisation |
| WIF1 | Wnt inhibitory factor 1 |
| Wnt | Wingless type |
| Yes | Yamaguchi sarcoma |

## Publications

**Heffron, CC**, Gallagher, MF, Guenther, S, Sherlock, J, Henfrey, R, Martin, C, Sheils, OM, and O'Leary, JJ (2007). Global mRNA Analysis to Determine a Transcriptome Profile of Cancer Stemness in a Mouse Model. Anticancer Research; 27, 1319-1324.

## Abstracts

Flavin, RJ, Gallagher, M, Elbaruni, S, **Heffron, CC**, O'Toole, S, Smyth, P, Laios, A, Aherne, S, Martin, C, Lao, K, Sheils, O, O'Leary, JJ (2009). Down regulation of miRNAs in Pluripotent Cancer Stem Cells is Mirrored in Advanced Ovarian Serous Carcinoma. Modern Pathology; 20, 213A.

Gallagher, MF, Elbaruni, S, **Heffron, CCBB**, Guenther, SM, Henfrey, R, Martin, C, Sheils, O, O'Leary JJ (2008). Characterisation of Early Stemness Regulation in Teratocarcinoma Stem Cells. Modern Pathology; 20, 327A.

Gallagher, MF, Elbaruni S, **Heffron, CCBB**, Guenther, SM, Henfrey, R, Martin, C, Sheils, O, O'Leary, JJ (2008). Differential Regulation of Key Stemness Genes During Early Differentiation of Teratocarcinoma Stem Cells. Modern Pathology; 20, 327A.

Gallagher, M, Elbaruni, S, **Heffron, C**, Guenther, S, Henfrey, R, Martin, C, Sheils, O, O'Leary, J (2008). Characterisation of Novel 'Early Stemness' Gene Events Specific to Highly-Malignant Cancer Stem Cells. Journal of Pathology; 214 (S1), 1A.

Gallagher, M, Elbaruni, S, **Heffron, C**, Guenther, S, Henfrey, R, Martin, C, Sheils, O, O'Leary, J. Early Regulation of 'Early Stemness' Gene Events in Pluripotent Cancer Stem Cells (2008). Journal of Pathology; 214(S1).

**Heffron CCBB**, Gallagher MF, Guenther S, Henfrey R, Sherlock J, Sheils O, Martin C, O'Leary JJ. Regulation of Cancer Stemness: Understanding Self-renewal and Early

Differentiation in Cancer Stem Cells. Platform Presentation, USCAP Annual Meeting, San Diego, USA, 24-30th March 2007.

**Heffron CCBB**, Gallagher MF, O'Leary JJ. Global mRNA Analysis for the Determination of a Transcriptome of Cancer Stemness in a Mouse Model. Poster Presentation, Stem Cell Symposium, University of Cambridge, UK, 18-19th December 2006.

**Heffron CCBB**, Gallagher MF, O'Leary JJ. Global mRNA Analysis for the Determination of a Transcriptome Profile of Cancer Stemness. Platform Presentation, NICB meeting, DCU, 7-8th September 2006 (Awarded Best Oral Presentation Award).

# CHAPTER 1 INTRODUCTION

## 1.1 Overview

The World Health Organisation (WHO) estimates that 7.6 million people died from cancer in 2005 with that number expected to increase to 9 million in 2015 and 11.5 million in 2030 (Organisation, 2007). A complete and comprehensive understanding of the molecular basis of the carcinogenic process is required to ultimately allow the successful prevention and treatment of cancer. While cancers in different organs and within organs may differ immensely, they also share a number of fundamental characteristics referred to by Hanahan and Weinberg, 2000 as the hallmarks of cancer (Hanahan et al., 2000):

1. Self sufficiency in growth signals
2. Insensitivity to growth inhibitory (anti-growth) signals
3. Evasions of programmed cell death (apoptosis)
4. Limitless replicative potential
5. Sustained angiogenesis
6. Tissue invasion and metastasis

The stem cell theory of cancer is a relatively old hypothesis but has been ignored by the prevailing assumption in the cancer field which assumes the stochastic model of carcinogenesis where transformation results from random mutations and subsequent clonal selection (Trosko et al., 2005; Tan et al., 2006). The concept of the existence of a 'malignant stem cell' as the cell of origin of tumours was initially proposed by Pierce as far back as 1974 (Pierce, 1974). The similarities between stem cells and cancer cells are striking and have been extensively documented in the literature. The defining features of a stem cell are its inherent abilities to both self renew and differentiate, features also present in cancer cells albeit in a less controlled manner. Cancer stem cells have been isolated in many tumours and most work has been done with haematological malignancies. Since then cancer stem cells have been demonstrated in prostate cancer (Collins et al., 2005; Collins et al., 2006), breast cancer (Al-Hajj et al., 2003), brain tumours (Singh et al., 2003), gastric cancer (Radtke et al., 2005), malignant melanoma (Klein et al., 2007), osteosarcoma (Gibbs et al., 2005), ovarian carcinoma (Bapat et al., 2005), lung cancer (Eramo et al., 2008) as well as many others. Conventional chemo and radiotherapies may fail due to the persistence of these cancer stem

cells post treatment thus resulting in local recurrence or metastasis. A recent study has identified a subset of embryonic stem (ES) cell-associated transcription regulators that are highly expressed in poorly differentiated tumours which have a poor clinical outcome (Ben-Porath et al., 2008). Applying our knowledge of the principles of stem cell biology could lead to a more extensive understanding of the regulation of cancer cell proliferation and progression and produce new targets for more effective cancer therapies. Successful elimination of the cancer stem cell, otherwise known as the cancer initiating cell, would be at the core of future treatments. In summary, the adoption of the cancer stem model of carcinogenesis will have a significant impact on the future management of cancer including its early detection, successful treatment as well as having important prognostic properties.

Breakdown in the regulation of self-renewal is thought likely to be the key event in the development of cancer (Clarke, 2005). Aberrant control of differentiation is another interpretation of what lies at the core of oncogenesis (Andrews et al., 2005). The understanding and elucidation of the pathways that control these stem cell properties is vital to the future of medical oncology.

### 1.2 Stem cells

The definition of a stem cell is a cell with the ability to differentiate into one or more specialised cell types, and the capacity for long-term self-renewal without senescence (Rippon, 2004). Stem cells can be totipotent, pluripotent or multipotent. Totipotent cells are cells capable of differentiating into every cell type including both embryonic and extraembryonic tissue, however, totipotency has proven difficult to achieve in vitro (Pan et al., 2002). The first critical step or decision in mammalian embryogenesis is the division of the embryo into two major lineages: the pluripotent inner cell mass (which generates all three germ cell layers) and the trophoblast (which supports embryonic growth). ES cells represent pluripotent stem cells and are derived from the inner cell mass of the blastocyst.

### 1.2.1 Pluripotent stem cells

A pluripotent stem cell is defined as a stem cell that has the potential to differentiate into tissues from the three germ cell layers, namely mesoderm, ectoderm and endoderm i.e. can produce all cell types except extraembryonic tissue (Ulloa-Montoya et al., 2005). Examples of pluripotent stem cells are ES and embryonal carcinoma (EC) cells. Indeed, much of our knowledge on pluripotent stem cells has come from research studies on non-human primate embryonic stem cells and embryonal carcinoma cells (Pal et al., 2006).

The pluripotent state is thought to be maintained in vitro by a number of transcription factors, the key factors being Oct4, Nanog and Sox2, all of which are expressed in the inner cell mass of the blastocyst and in undifferentiated ES cells (Yamanaka et al., 2008). Oct4 and Nanog are homeodomain proteins and Sox2 belongs to a group of SRY (sex determining region Y)-related HMG (high mobility group)-box (Sox)-containing proteins.

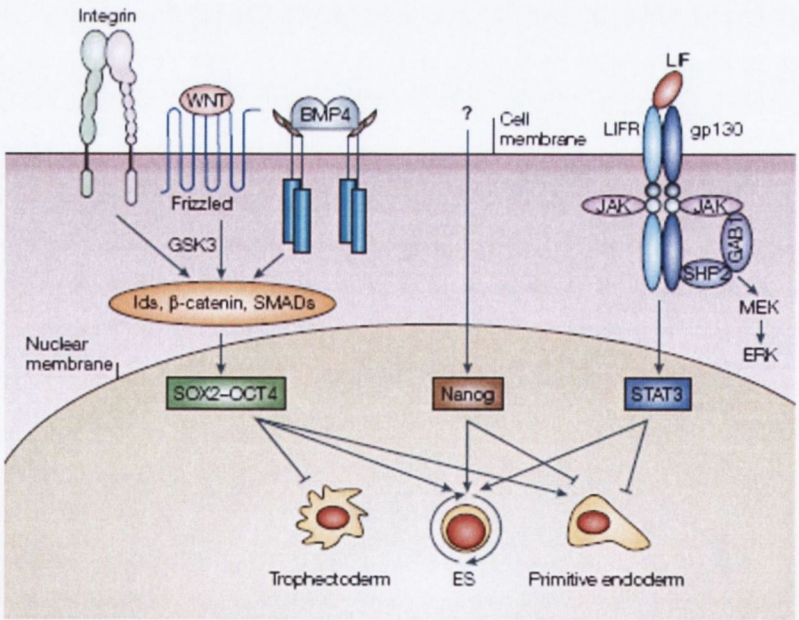



**Figure 1.1: Roles of Oct4, Sox2 and Nanog in the signalling pathways involved in maintaining pluripotency in ES cells (Boiani et al., 2005). Cell-surface receptors initiate signals that are communicated to the nucleus and affect the key transcription regulators, namely Oct4, Sox2 and Nanog. The question mark refers the current lack of information regarding the surface receptors and transducers involved in the regulation of Nanog.**

Oct4 is a transcription factor that belongs to the POU transcription factor family. The acronym POU is derived from the names of three mammalian transcription factors, namely the pituitary-specific Pit-1, the octamer transcription factor proteins (Oct1 and Oct2) and the neural Unc-86 transcription factor from Caenorhabditis elegans. Its expression is restricted to pluripotent cells and it is essential for the initial development of pluripotentiality in the inner cell mass. Mouse embryos with mutated Oct4 do not survive implantation. The precise level of Oct4 is important for determining ES cell fates – too much results in differentiation into mesoderm and primitive endoderm whereas too little induces the formation of trophoectoderm and thus a loss of pluripotency. Forced constitutive expression of Oct4 cannot prevent differentiation of ES cells and a less than 2 fold increase in expression actually causes differentiation into primitive endoderm and mesoderm. Thus a critical amount of Oct4 is required to sustain stem cell self-renewal but is not sufficient to prevent differentiation (Palmqvist et al., 2005). Oct4 is thought to work in conjunction with another transcription factor, namely Nanog, to modulate transcription programmes necessary for the determination of ES cell fate (Liang et al., 2008).

Nanog is another transcription factor and homeodomain protein also being a key regulator of pluripotentiality. It was named after Tír Na nÓg, the mythological Celtic land of the 'ever young' (Boiani et al., 2005). While it is found in ES cells, it is largely absent from differentiated cells. The levels of Nanog expression do not appear to be as critical as those of Oct4 with elevated levels of Nanog conferring constitutive self-renewal of mouse ES (mES) cells and allowing them to grow in the absence of additional factors such as leukaemia inhibitory factor (LIF). A similar finding has been demonstrated with human ES (hES) cells, with over expression of Nanog enabling cells to grow without feeder cells (Darr et al., 2006). It belongs to the homeobox set of genes and is undetectable in most normal human tissues but highly expressed in ES cells.

Sox transcription factors, of which there are 20 identified in the mouse, are crucial for embryonic development and play critical roles in cell fate determination, differentiation and proliferation. Human Sox2 shows 98% homology with mouse Sox2 and it is highly conserved between vertebrates. Sox2 is an essential factor in the maintenance of pluripotency and it has been shown that Sox2 is necessary to regulate multiple transcription factors that affect Oct4 expression (Masui et al., 2007). Sox2 forms a ternary complex with Oct4 enabling it to

participate in the regulation of the inner cell mass and its derivative cells. A reduction of Sox2 expression will induce mES cells to differentiate into trophoectoderm. The role of Sox2 in gastric carcinogenesis has been a subject of a recent study by Otsubo et al, 2008, who found that Sox2 expression was reduced in gastric carcinomas compared with normal gastric epithelial cells. They suggest that Sox2 may control the gastric epithelial cells differentiating into mature cells and that its disruption may cause continual dividing, thus eventually leading to gastric cancer (Otsubo et al., 2008).

Signal transducer and activator of transcription-3 (Stat3) and the glycoprotein-130 (gp130) receptor-mediated pathway via the LIF signal cascade plays an important role in mES cell self-renewal and differentiation. LIF exerts its effect by binding to the LIF receptor-gp130 heterodimer receptor on the cell membrane and activates Stat3 as shown in Figure 1.1. Six key Stats (1-6) have been identified. On activation, Stat3 becomes phosphorylated and translocates to the nucleus where it acts as a transcription factor. In addition, nonreceptor tyrosine kinase Janus (JAK) and the antiphosphotyrosine immunoreactive kinase (TIK) are recruited thus activating other pathways. The intracellular second messengers, PI3K (phosphoinositide 3-kinase) products, have also been shown to be activated by the gp130 component of the LIF receptor playing a role in the regulation of self-renewal of mES cells (Paling et al., 2004). The LIF-STAT3 signalling pathway has been shown to support the self-renewal of mES cells while in hES cells it does not appear to prevent differentiation (Sato et al., 2004).

Transforming growth factor (TGF)-β and TGF-β related proteins, such as the bone morphogenetic proteins (BMPs), are yet another group of signalling molecules important in stem cell biology. They act in maintaining the undifferentiated state as well as in the selection of a differentiation lineage and progression of differentiation along an individual lineage. When TGF-β signalling is disrupted, cancer may ensue due to impaired differentiation. The TGF-β family consists of about 30 structurally related growth and differentiation factors including TGF-βs, activins, nodal and BMPs (Mishra et al., 2005). These factors employ transmembrane receptor serine-threonine kinases to activate Smad (Similar to mothers against decapentaplegic homologue) transcription factors resulting in their nuclear relocation and subsequent activation or repression of gene expression. Smads are transcription factors that are utilised by a number of signalling pathways including Wnt (Wingless type) and Hedgehog signalling pathways.

### *1.2.1.1 Embryonic stem cell*

Embryonic stem cells are derived from the inner cell mass (ICM) of mammalian blastocysts and maintain pluripotency, an ability to differentiate into all types of somatic and germ cells as illustrated in Figure 1.2. Only the cells present in the fertilised egg or in the zygote until the 8 cell stage retain the capacity to generate an entire organism with the required embryonic and extraembryonic tissues and thus are the only true totipotent cells. Cells from the inner cell mass are considered pluripotent as they retain the ability to develop only into cells found in the embryo proper.

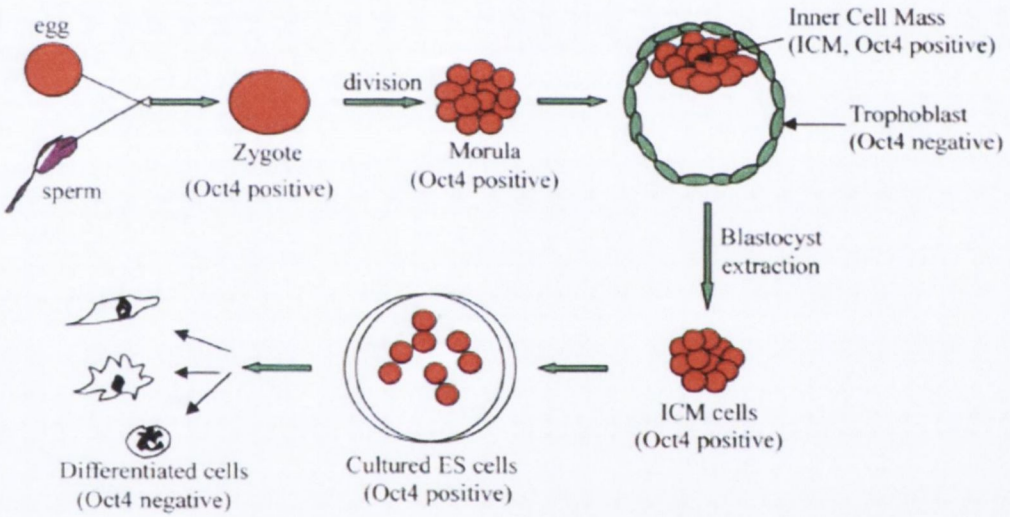



**Figure 1.2: Origin of ES cells (Pan et al., 2002). The isolation and differentiation of ES cells in vitro are illustrated schematically in this diagram starting with the fertilisation of an egg by a sperm to form a zygote. At the blastocyst stage, the inner cell mass (ICM) becomes visible and this is the source of ES cells. These ES cells can then be cultured to produce various cell types. The presence or absence of Oct4 expression at each stage is also noted.**

The derivation of mES cell lines was reported for the first time in 1981 by Martin and Evans et al and have proven very useful in the study of mammalian development (Evans et al., 1981; Martin, 1981). Indeed, they heralded a major breakthrough for development biology as they

provided the means to study early embryonic development as well as cellular differentiation (Martin et al., 1975). Over the past 20 years, they have played a major part in shaping the future of medicine providing information on drug development, directed differentiation as a method to treat disease, nuclear transfer protocols in cloning and also in the establishment of methodologies for the isolation of non-rodent ES cells (Downing et al., 2004). A widely studied example of directed differentiation is in the development of a potential treatment for Parkinson's disease in which the formation of dopaminergic and serotonergic neurons from mES cells may provide the required information for the practical use of cell replacement therapy in Parkinson's disease (Sasai, 2002). Mouse ES cells can be maintained in culture indefinitely without loss of their broad pluripotent capacity as determined by their ability to differentiate and give rise to all three germ cell layers both in vitro and in vivo.

Pluripotency of ES cells has been demonstrated by the following three properties (Itskovitz-Eldor et al., 2000):

1. ES cells injected into blastocysts of pregnant mice contribute to all cell types in the chimeric progeny mice, including the germ layer.
2. ES cells injected subcutaneously into syngeneic mice induce teratoma formation (these tumours may include cells from endodermal, mesodermal and ectodermal origin).
3. In vitro aggregation of ES cells results in formation of embryoid bodies with regional differentiation into embryonically distinct cell types.

Human ES cell lines were successfully generated almost twenty years later in 1998 by Thomson et al thus giving rise to hope in the field of regenerative medicine and tissue engineering and a real possibility of significant advances in the treatment of human disease (Thomson et al., 1998). They isolated five cell lines originating from five separate embryos using similar methods described for non human primate ES cells with the pluripotent properties of ES cells as outlined above. Their work was later replicated and confirmed by other groups (Reubinoff et al., 2000). Human ES cell lines were found to express cell surface markers that also characterise undifferentiated non human primate ES and human EC cells, including stage-specific embryonic antigen (SSEA)-3, SSEA-4, TRA-1-60, TRA-1-81, and alkaline phosphatase (Badcock et al., 1999). In comparison, mES cells express SSEA-1 and not SSEA-3 or SSEA-4.

The use of hES cells for research purposes creates a major ethical dilemma and in many countries including Ireland, their use in experimental research is not permitted. Therefore, research using mES cells remains a vital source of results in the field of stem cell biology. The following is a summary of some of the uses and functions of ES cells:

- To identify new genes involved in promoting and inhibiting differentiation.
- As cellular biological models of disease.
- Used to introduce subtle modifications to the nuclear genome in gene targeting experiments.
- Used in directed differentiation methodologies for specific diseases.

#### *1.2.1.2 Embryonal carcinoma cells*

The first pluripotent cell lines to be established were embryonal carcinoma (EC) cell lines which were derived from the undifferentiated compartment of murine germ cell tumours in 1967 by Finch et al and again in 1975 by Martin et al (Finch et al., 1967; Martin et al., 1975). Indeed, the development of these cell lines preceded the development of ES cells and much of the initial work in cell differentiation and development was performed on these. EC cell lines provided a practical in vitro model of differentiation and produced a vast quantity of information on cell differentiation and early embryogenesis.

**Table 1.1: Key cell lines used in mES cell line work over the last 20 years (*refers to embryonal carcinoma cell lines).**

| Cell line | Year established |
|---|---|
| F9* | 1973 |
| P19* | 1982 |
| CCE | 1981 |
| D3 | 1985 |
| E14 lines | 1987 |
| AB lines | 1990 |
| J1 | 1992 |
| R1 | 1993 |

EC cells represent the stem cells of teratocarcinoma (Astigiano et al., 2005) and are widely considered to be the malignant equivalent of ES stem cells whether human or mouse (Andrews, 2002; Andrews et al., 2005; Pal et al., 2006). Figure 1.3 illustrates the relationship between ES and EC cells.

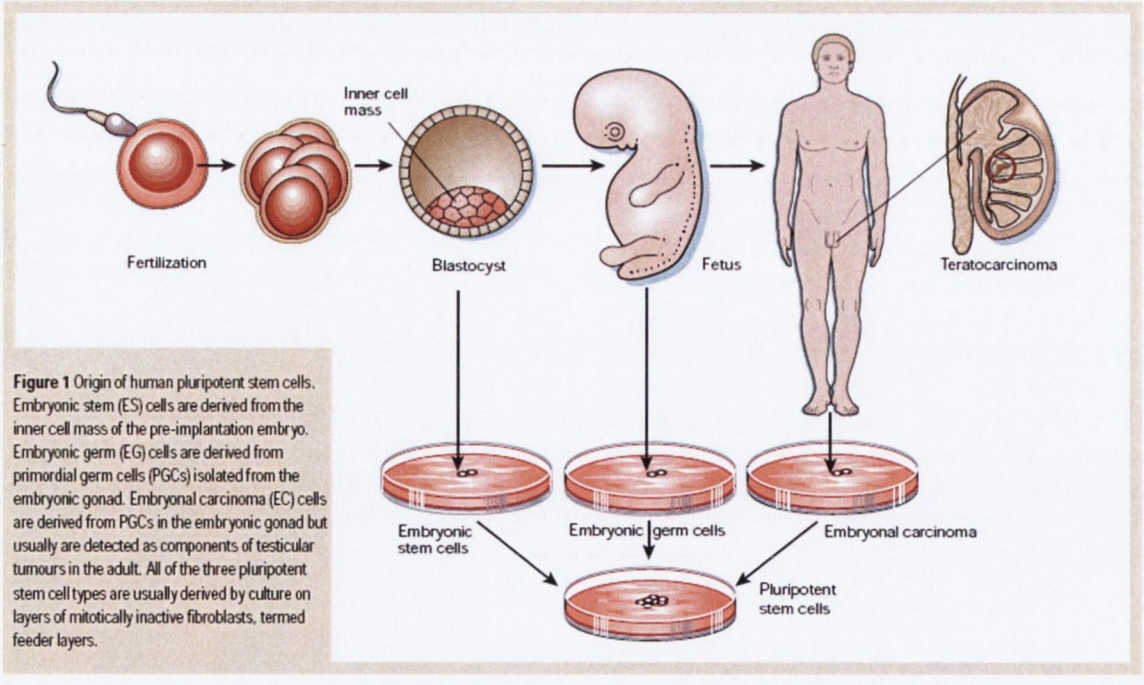



**Figure 1.3: Illustration of the different types of pluripotent stem cells, namely embryonic stem cells and embryonal carcinoma cells (Donovan et al., 2001).**

As many of the molecular pathways that underlie tumourigenesis are also involved in normal embryogenesis, the study of EC cells should provide an excellent model to study the mechanisms involved in tumour growth. Currently, they are widely used as a model system in the laboratory in the study of embryonic development and differentiation. Given the limitations in use of hES cells, human EC cells remain a useful and viable alternative to further our knowledge and understanding of pluripotency, cell differentiation as well as early embryogenesis. However, being cancer-derived and usually aneuploid, EC cells are not always useful for clinical application (Rippon, 2004).

Nullipotent EC cell lines have also been developed for both mouse and human. Nullipotent refers to the inability of the EC cells to differentiate into other cell types as their pluripotent counterpart does. A comparison of the two tumour types showed that the injection of pluripotent EC cells into nude mice produced teratomas while the injection of nullipotent EC cells produced a pure embryonal carcinoma (Duran et al., 2001). Nullipotent EC cell lines are easy to maintain and strongly resistant to differentiation stimuli.

### 1.2.2 Other types of stem cells

*Somatic stem cell:* Somatic stem cells are responsible for normal tissue renewal (Gudjonsson et al., 2005). They are thought to be multipotent while ES cells are pluripotent. Multipotent refers to their limited ability to differentiate as multipotent cells only differentiate into cells that make up the given organ they are associated with unlike pluripotent cells which are capable of differentiating into cells of the mesoderm, ectoderm and endoderm. Haemopoietic stem cells, epithelial stem cells in the lining of the digestive tract and skin stem cells are examples. Alternatively, they can be referred to as adult stem cells. They have been identified in many organs and tissues and reside in a specific area of each tissue where they remain quiescent until required following tissue injury or disease.

However, more recently, some adult stem cells have been shown to have some pluripotent properties. Mesenchymal stem cells found in the bone marrow are adult stem cells and were found to be capable of differentiating into lung epithelial cells as well as their usual role in

producing cells within the bone marrow (Wang et al., 2005a). Another study has shown the production of gastrointestinal epithelial cells from bone marrow derived cells (Matsumoto et al., 2005).

### 1.2.3 Differentiation and self-renewal in stem cells

Differentiation and self-renewal are the defining features of stem cells. The cancer stem cell theory suggests that aberrant control of differentiation or self-renewal lies at the core of oncogenesis (Andrews et al., 2005). EC cells from teratocarcinomas and ES cells from embryos are regarded as different points along the same spectrum and offer a unique model to both study and explore differentiation and self-renewal in stem cells and potentially the mechanisms of carcinogenesis.

Studies using mES cells show that a set of transcription factors is important for the maintenance of self-renewal and the inhibition of differentiation with Oct4, Nanog and Sox2 as described previously in section 1.2.1 being the main players. A number of other factors have been shown to be involved in differentiation including fibroblast growth factors (FGFs) and BMPs which appear to work in conjunction with each other and other transcription factors in the differentiation pathway (O'Shea, 2004). Nodal and Wnt signalling are thought to have roles in the early embryo development and their activation is required together for the formation of primitive streak like cells in ES cell differentiation cultures (Gadue et al., 2006). The primitive streak is a structure that forms during the early stages of mammalian embryo development. Within different areas of the primitive streak, meso, endo and ectoderm formation are induced, i.e. the formation of the three germ cell layers. Other signalling pathways important for and involved in self-renewal and differentiation are sonic hedgehog (SHH), BMI1, Notch and Phosphatase and tensin homologue deleted from chromosome 10 (PTEN) signalling pathways (Park et al., 2004).

ES cells grown in vitro form small aggregates of cells surrounded by an outer layer of visceral endoderm and these aggregates have been termed embryoid bodies (EBs) (Keller, 1995). These EBs are an important source of information on differentiation and gene expression in early development. This pattern of ES development as EBs parallels that of the early embryo.

As EBs grow and differentiate, they recapitulate the normal embryo and contain differentiated tissues with pulsating cardiomyocytes and neurons some of the first cells to develop.

### 1.2.4 Cancer stem cells (CSC)

The existence and concept of cancer stem cells are disputed by many despite their existence being first proposed over 40 years ago. The best evidence supporting their existence comes from the study of haematological malignancies in which a small subset of cells is thought to be responsible for the neoplasm (Lapidot et al., 1994; Bonnet et al., 1997; Wang et al., 2005b). Within the haematological malignancies, chronic myeloid leukaemia (CML) and acute myeloid leukaemia (AML) have served as important model diseases in the discussion of cancer stem cells.

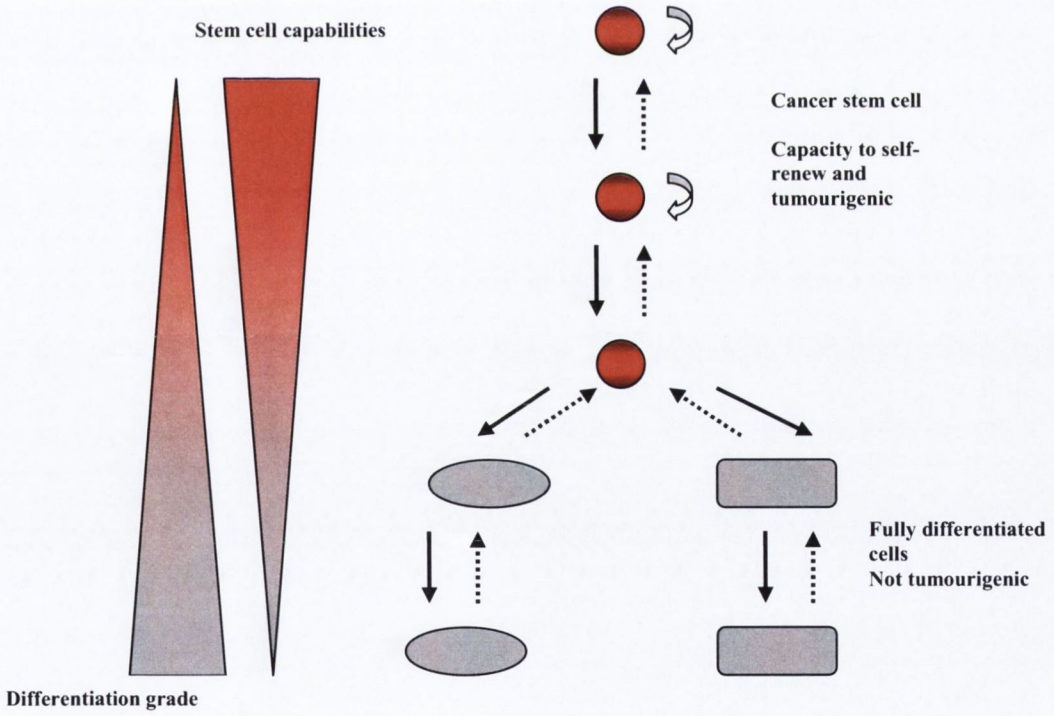



**Figure 1.4: Illustration of the CSC model with the CSC on the top of the hierarchy (red) with the ability to self-renew, i.e. generating more CSCs, and to produce more differentiated cells (grey) (Vermeulen et al., 2008).**

To date, their existence has also been shown in breast tumours (Al-Hajj et al., 2003), brain tumours (Singh et al., 2003; Singh et al., 2004), osteosarcomas (Gibbs et al., 2005), malignant melanoma and prostate tumours (Collins et al., 2005), gastric tumours, ovarian tumours (Bapat et al., 2005), lung cancer (Eramo et al., 2008), to name but a few. A German group found the expression of stem cell markers in soft tissue sarcomas to be a poor prognostic factor thus suggesting that a tumour with a stem cell signature will have significantly poorer prognosis (Taubert et al., 2007)

A number of definitions for a cancer stem cell have been offered by various authors. Smith, 2006 stated that a cancer stem cell could be defined as a 'self-renewing cell responsible for sustaining a cancer and for producing differentiated progeny that form the bulk of the cancer' (Smith, 2006). He also went on to say that 'cancer stem cells identified in leukaemias and certain solid tumours are critical therapeutic targets'. Bjerkvig et al, 2005 refer to a cancer stem cell being an operational term and define a cancer stem cell as 'a cancer cell that has the ability to self renew; dividing to give rise to another malignant stem cell and a cell that gives rise to the phenotypically diverse tumour cell population' (Bjerkvig et al., 2005). Some authors view cancer as a stem cell disease (Gudjonsson et al., 2005). Essentially, cancer stem cells can be referred to as a subset of cancer cells capable of self-renewal to generate additional cancer stem cells and also possessing the ability to differentiate into a diverse population of differentiated cancer cells with limited proliferative ability and are sometimes referred to as cancer initiating cells.

The hypothesis is that malignant tumours are comprised of both cancer stem cells, which have great proliferative potential, as well as more differentiated cancer cells with limited proliferative potential which form the bulk of the tumour. Thus, these two types of tumour cells would be phenotypically different and respond differently to chemotherapies. This is demonstrated in the haematological malignancy, CML, which possesses a defining cytogenetic abnormality known as the Philadelphia chromosome. The chromosomal abnormality is a translocation between chromosomes 9 and 22 resulting in a BCR-ABL fusion protein. This has lead to the production of a specific BCR-ABL inhibitor, imatinib, which targets the bulk of the tumour or differentiated component. Resistance to imatinib is thought to be a result of the presence residual tumour cells or cancer stem cells which are resistant to imatinib thus

providing further evidence of the existence of the cancer stem cell (Wang et al., 2005b; Mauro, 2006; Melo et al., 2007; Savona et al., 2008). The biology of teratocarcinomas, malignant tumours that occur in the ovary and testes of humans composed of undifferentiated embryonal carcinoma cells and a differentiated teratomatous component, are thought to reflect this model.

### *1.2.4.1 Cellular origin of cancer stem cells*

The cellular origin of cancer stem cells has not been definitively determined. The focus of much cancer research is to determine the cell from which the tumour originates thus gaining insight into the first steps of the neoplastic process. Some consider the cancer stem cell as the biologically distinct cell within a neoplastic clone that is capable of initiating and sustaining tumour growth, i.e. the cancer-initiating cell, with the cell of origin referring to the normal cell in which the initial transforming event occurs thus making the concept of a cancer stem cell a more functional rather than absolute term. Therefore, a cancer stem cell is not necessarily derived from a normal tissue stem cell, i.e. the genetic cancer causing mutation does not always occur in the normal stem cell (Tan et al., 2006; Vermeulen et al., 2008). Thus, regardless of the cell of origin, the cancer stem cell is defined by its stem cell-like properties.

The principal current theories for the origin of the cancer stem cell include:

- Mutational transformation of normal stem cells.
    - Due to their longevity and specific self-renewing properties, it is believed they have a greater propensity to accumulate carcinogenic mutations compared with short-lived, differentiated cells and thus are an ideal target of the carcinogenic process (Gudjonsson et al., 2005).
    - Within AML, the number of similarities between the leukaemic stem cells and the haematopoietic stem cells suggests that the initial carcinogenic event is occurring in the stem cell rather than in the committed cell progenitors (Wang et al., 2005b).
- Cell fusion.
- Horizontal gene transfer processes.

- Imbalance between symmetric and asymmetric cell division.

### *1.2.4.2 Pathways involved in cancer stem cells*

Mutations in the pathways that regulate self-renewal in normal stem cells have been found in cancer cells. For example, mutations and dysregulation of the Wnt signalling pathway, a pathway known to be involved in the regulation of self-renewal in normal stem cells, have been found in colorectal cancers and leukaemias. In fact, many authors believe that dysregulation of self-renewal is the key event in carcinogenesis (Wicha et al., 2006). The pathways involved in the regulation of normal stem cells such as Wnt, Notch and Hedgehog have been shown to be dysregulated in many cancers, illustrated in Figure 1.5.

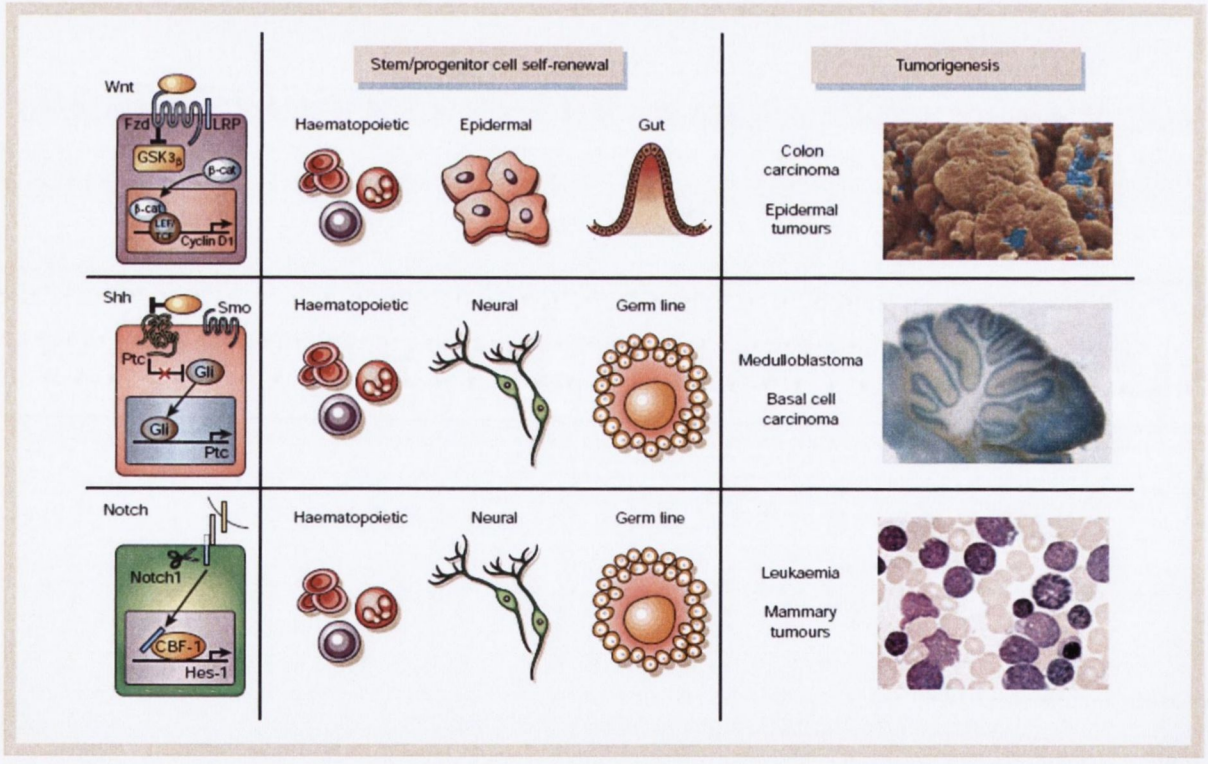



**Figure 1.5: This figure shows examples of pathways involved in self-renewal regulation and then in various tumours (Reya et al., 2001).**

**Table 1.2: Signalling pathways, stem cells and cancer (Pardal et al., 2003).**

| Pathway | Stem Cell | Cancer |
|---|---|---|
| ***Wnt*** | Haematopoietic stem cells | Lymphoblastic leukaemia |
| | Intestinal stem cells | Colorectal cancer |
| | Keratinocyte stem cells | Pilomatricoma |
| | Cerebellar granule-cell progenitors | Medulloblastoma |
| | CNS stem cells | |
| ***SHH*** | Hair-follicle progenitors | Basal cell carcinoma |
| | Cerebellar granule-cell progenitors | Medulloblastoma |
| | CNS stem cells | Gliomas |
| ***BMI1*** | Haematopoietic stem cells | B cell lymphomas |
| | | AML |
| ***Notch*** | Haematopoietic stem cell | Lymphoblastic leukaemia |
| | Mammary epithelial stem cells | Breast cancer |
| ***PTEN*** | Neural stem cells | Gliomas |

The differences between signalling pathways in normal and cancer stem cells have yet to be definitively elucidated (Zhang et al., 2006). The eventual goal of such knowledge will be to design new therapeutic targets with the eventual purpose of eliminating any residual tumour cells and thus preventing recurrence.

The Wnt signalling pathway along with other signalling pathways was thought to be the major molecular mechanism behind embryonic development, playing a major role in the maintenance of the self-renewal state in both mES and hES cells. Now many of the genes in the Wnt pathway are also thought to act as oncogenes and tumour suppressor genes when deregulated in cancer (Klaus et al., 2008). Some of the components of the Wnt signalling pathway include the Wnt proteins (19 Wnt genes in the human genome), Frizzled receptors, LRP5 and 6 co-receptors as well as Frizzled-related proteins (SFRPs), Dickkoptfs (DKKs) and Wnt inhibitory factor 1 (WIF1). The protooncogene, MYC, was identified as a direct transcriptional target of the Wnt-β-catenin signalling pathway thus highlighting the involvement of the Wnt signalling pathway in cancer development. Since the inappropriate activation of the Wnt pathway was first linked to human cancer, researchers and

biotechnology companies have been interested in developing Wnt pathway inhibitors. Wnt signalling has also been shown to have an important role in the stem cell compartments of various tissues and appears to join forces with Notch signalling. Wnt-β-catenin signalling has been shown to have an essential function in the maintenance of the mammary gland and skin cancer stem cells giving rise to the possibility of exploitation for therapy with Wnt inhibiting drugs thus targeting the putative cancer stem cells which are often resistant to other forms of treatment (Malanchi et al., 2008). The Wnt-β-catenin signalling pathway has also been shown to be activated further in the late stages of tumour progression in colonic tumours, therefore, Wnt inhibiting drugs may also interfere with metastasis formation and maintenance (Stein et al., 2006).

#### *1.2.4.3 Identifying cancer stem cells*

As cancer stem cells may only account for < 0.1% of a given tumour, it can be very difficult to isolate these cells for further study. Research in this area has been hindered by a lack of methods to isolate subpopulations of cells from tissue samples. Efforts to rectify this are under way with the development of marker profiles and identification of cell surface markers. Cell surface markers such as CD133 and CD44 have been shown to be expressed in small subsets of cells in brain (Singh et al., 2004), prostate (Richardson et al., 2004; Collins et al., 2005) and breast (Al-Hajj et al., 2003) tumours which appear to account for all the in vitro proliferative activity and further tumour production. Thus, it is proposed that such markers may be used to prospectively identify cancer stem cells. Many groups are using these markers to successfully isolate the cancer stem population with some interesting results. The cancer stem cell isolate in prostate cancer has been found to be androgen resistant thus explaining its resistance to androgen prostate cancer treatments resulting in treatment failure. Bonnet and Dick demonstrated a CD34 positive cell population within AML which were the leukaemia-initiating cells and produced AML in nude mice (Bonnet et al., 1997). However, further characterisation of cancer stem cells may uncover more straightforward and efficient methods of identification.

### *1.2.4.4 Cancer stem cells and potential therapies*

Targeting cancer stem cells represents a potentially effective means of inhibiting tumour progression and disease recurrence, but this approach does require the ability to block the self-renewal capability of a small population of cancer stem cells without toxicity to the normal stem cell population (Samuel et al., 2005). A variety of stem cells have been shown to express drug–resistance proteins which would make them more resistant to treatment with conventional chemotherapeutic agents. They are known to be slow cell cyclers and express high levels of drug transporters conferring drug resistant properties to the cancer stem cells (Lou et al., 2007). Novel gene expression profiling experiments have shown the existence of androgen resistant genes in a population of cancer stem cells isolated from prostate tumours (Collins et al., 2005). This has been proposed to offer an explanation for the failure of chemotherapeutic regimes which may only shrink the tumour rather than also eliminating the cancer stem cells. Thus these cells would remain to continue the process of tumour growth and differentiation. One group found that within glioblastoma, a lethal primary brain tumour with a very poor prognosis due to radioresistance, a population of CD133 positive tumour cells was responsible for conferring glioma radioresistance and thus may be the source of tumour recurrence following radiation treatment (Bao et al., 2006).

The identification of agents that would specifically target the cancer stem cell population in any given tumours would greatly enhance the success of a given treatment.

1. Specific targeting of cancer stem cells:

Uncovering the essential differences between the normal stem cell and the cancer stem cell may reveal new targets for future treatments in an attempt to eliminate the offending cancer stem cell and thus prevent tumour recurrence and metastasis as illustrated in Figure 1.6.

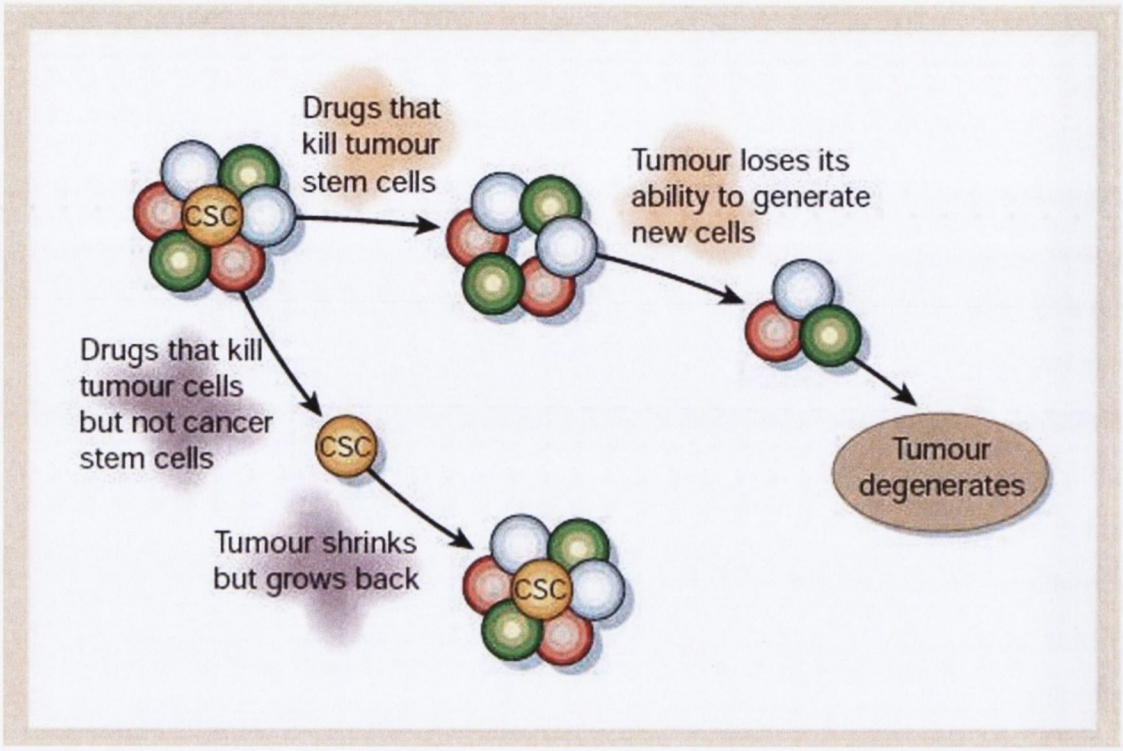



**Figure 1.6: Treatment of the cancer stem cell (Reya et al., 2001). Conventional therapies may shrink the tumour, killing cells with limited proliferative potential with the cancer stem cells remaining viable. On the other hand, targeted cancer stem cell therapies might more effectively kill the cancer stem cells, rendering the tumours unable to maintain themselves or grow.**

One example has been identified in colon cancer where it was found that interleukin (IL)-4 conferred survival advantages to colon carcinoma stem cells thus allowing them to resist standard chemotherapies (Todaro et al., 2007). The authors also found that neutralisation of IL-4 targeted this stem cell population achieving the successful elimination of the cancer initiating cells, and suggest the use of anti-IL-4 may be a very promising finding in the search for successful cancer treatments.

The question of therapies directed towards cancer stem cells affecting normal stem cells as the same pathways involved in the survival and neoplastic proliferation of cancer stem cells also maintain normal stem cell populations may arise. It appears that cancer stem cells are likely to be more dependent on some pathways than normal stem cells thus therapies targeting particular pathways may be more useful. In other cases, it may be possible to do without the

normal stem cell (Pardal et al., 2003). For example, in breast cancer, the elimination of mammary epithelial stem cells may be acceptable as it is also lost in patients who undergo a mastectomy. However, ultimately, it is an important consideration in the development of new stem cell targeted therapies.

2. Therapies to induce differentiation of stem cells or even transiently inhibit the maintenance of the stem cell state:

This group of therapies should lead to the exhaustion of the pool of cancer stem cells and to the conversion of malignant cancers into benign tumours. For example, transient inactivation of MYC leads to the differentiation of sarcoma cells into osteocytes and a loss of neoplastic phenotype that cannot be restored even by reactivation of MYC. This raises the possibility that MYC is required for maintenance of identity of sarcoma stem cells, such that in its absence the cells differentiate to benign osteocytes (Gibbs et al., 2005). Another example is the model of embryonal carcinoma used in this study, where despite the aggressive nature of embryonal carcinoma cells, they can be induced to differentiate and subsequently lose their tumourigenic potential (Astigiano et al., 2005). Interestingly, embryonal carcinoma cells injected into a mouse blastocyst have been shown to become incorporated into the developing embryo indicating that there are external microenviromental factors involved (Astigiano et al., 2005).

3. Targeting the stem cell niche:

The microenvironment of stem cells or the so called 'stem cell niche' contains many cellular and non cellular components including extracellular matrix proteins that may be potential sources of therapies (see section 1.2.4.5 for a detailed discussion).

Figure 1.7 illustrates an overview of the cancer stem cell model and the effect of clinical interventions on cancer stem cells.

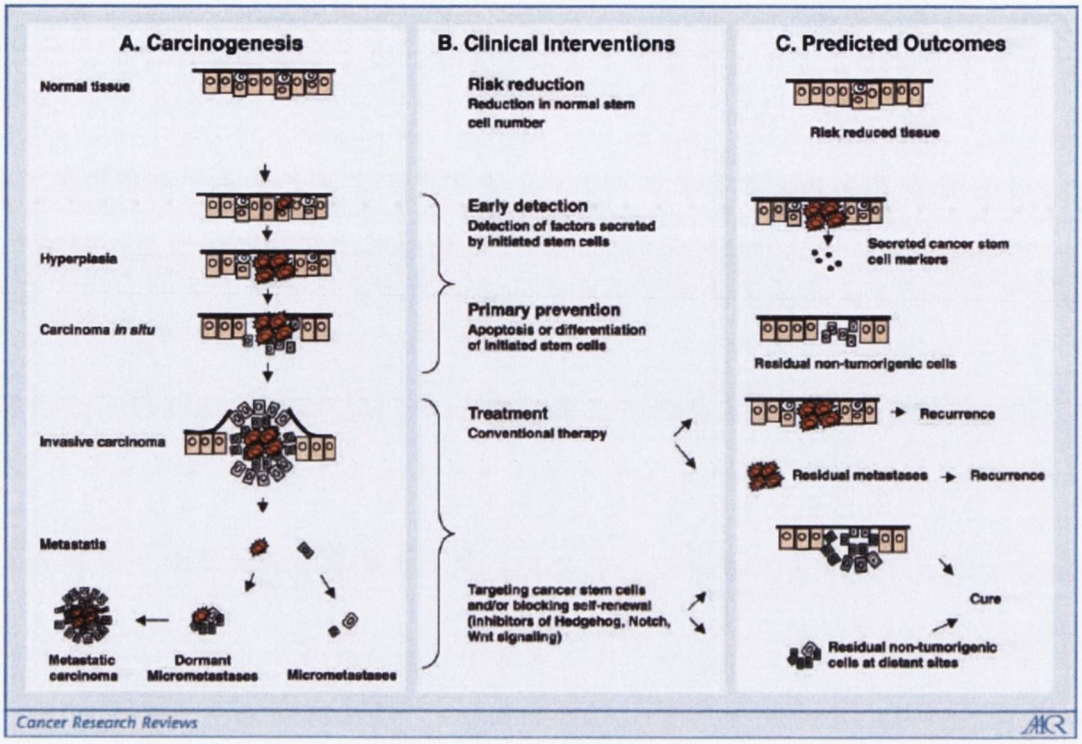



**Figure 1.7: Clinical implications of the cancer stem cell model (Wicha et al., 2006). It has important implications for clinical intervention as illustrated.**

Interestingly, Wicha et al suggest that the cancer stem cell hypothesis indicates that molecular profiling of tumours may not be as useful as expected in predicting sensitivity to therapy as these treatments would target the differentiated population of cells accounting for the bulk of the tumour and not the cancer stem cells. Therefore, specific profiling of the cancer stem cells would be the most beneficial way of determining new effective therapies.

### *1.2.4.5 Microenvironment of stem cells/Stem cell niche*

It is generally well recognised that the microenvironment of stem cells provides a fundamental role in their growth and development. The concept of the stem cell niche was developed by Schofield in 1978 who proposed that stem cells reside within a fixed compartment or niche designed to maintain stem cells (Schofield, 1978). Secreted and cell surface molecules are released into this defined anatomical compartment and control the rate of stem cell proliferation, determine the fate of stem cells and protect stem cells from death. Many of these

molecules are Wnt molecules, BMPs, FGFs, and Notch which are all known to play roles in controlling stem cell self-renewal as well as regulating lineage fate in different systems (Li et al., 2006). Subsequent studies have provided evidence to support Schofield's hypothesis with much of the work being performed on haemopoietic stem cells (Rizo et al., 2006).

The components of the stem cell niche are as follows:

1. The stem cell itself.
2. Stromal support cells.
3. Extracellular matrix proteins.
4. Blood vessels.
5. Neural inputs.

The features of stem cell niches as outlined by a number of groups are as follows (Angstreich et al., 2005; Jones et al., 2008):

1. Signals produced by the niche regulated stem cell self-renewal, survival and maintenance.
2. The spatial relationship between stem cells and support cells can polarise stem cells within the niche to promote asymmetric stem cell divisions.
3. Adhesion between stem cells and supporting stromal cells and/or the extracellular matrix (ECM) anchors stem cells within the niche in close proximity to self-renewal and survival signals.

Within the field of tumourigenesis where cancers are being considered as diseases of self-renewing cancer stem cells, the microenvironment or niche may play a vital role in the development of a tumour. Various suggestions have been made regarding how the microenvironment contributes with some suggesting an independence from niche requirements and others suggesting that the niche may be involved directly in either tumour initiation and/or tumour maintenance. Sneddon et al, 2006, performed cDNA microarray analysis on stromal cells from basal cell carcinomas (BCC) and non tumour skin and found that BMP antagonists secreted by stromal cells in BCC allowed continued proliferation and

self-renewal of cancer cells, thus providing evidence that the tumour microenvironment plays a role in tumour development (Sneddon et al., 2006).

Another interesting example of the effect of the microenvironment is provided in a study performed by Astigiano et al., 2005, in which they showed that the post-implantation mouse embryo microenvironment retained an ability to control cancer growth by allowing embryonal carcinoma cells to be incorporated into normal embryogenesis if transplanted early and forming tumours if transplanted later in development (Astigiano et al., 2005). The mechanisms involved have not yet been identified but the authors suggest that epigenetic mechanisms and external signals are involved rather than any genetic changes.

The presence of a stem cell niche and tumour cell niche provides a possible target for the development of new cancer treatments as well as being of significance in other stem cell areas such as regenerative medicine and stem cell transplants. One study demonstrated the pharmacologic use of parathyroid hormone increased the number of haemopoietic stem cells mobilised into the blood for stem cell harvests, protected stem cells from repeated exposure to cytotoxic chemotherapy and expanded stem cells for transplant recipients (Adams et al., 2007). Thus, the stem cell niche may provide a useful and attractive target for drug based stem cell therapeutics. For tumours, directly targeting the tumour cell niche or preventing the tumour cell access to the niche may reverse or delay tumour progression. Jin et al 2006 found that targeting CD44, a molecule that mediates adhesive cell to cell and cell to extracellular matrix interactions eradicated human AML leukaemic stem cells (LSC) (Jin et al., 2006). These results provide evidence that the tumour cell niche or microenvironment may be a valuable source of future cancer therapies.

### 1.2.5 Current research on stem cells

A seminal article by Ramalho-Santos et al was published in Science in 2002 on stem cells and introduced the concept of 'stemness' (Ramalho-Santos et al., 2002). They compared the transcriptional profiles of mouse embryonic, neural, and haematopoietic stem cells as well as differentiated cells from lateral ventricles and the bone marrow to define a genetic program for stem cells. Subsequent gene lists were compared with each other to determine an overlapping

set of genes. This analysis resulted in a list of 216 genes which were enriched in each of the stem cell types, i.e. mouse embryonic, neural and haematopoietic stem cells. Based on their findings, they proposed that the essential attributes of stemness included:

1. Active JAK/STAT (Janus kinase/signal transducers and activators of transcription), TGF-β, Yes (Yamaguchi sarcoma) kinase and Notch signalling.
2. Capacity to sense growth hormone and thrombin.
3. Interaction with the extracellular matrix via integrin alpha6/Beta1, Adam9 and bystin.
4. Engagement in the cell cycle, either arrested in G1 or cycling.
5. High resistance to stress with upregulated DNA repair, protein folding, ubiquitin system and detoxifier system.
6. A remodelled chromatin acted upon by DNA helicases, DNA methylases and histone deacetylases.
7. Translation regulated by RNA helicases of the Vasa type.

## 1.3 Embryonal carcinoma, teratocarcinoma and teratomas

### 1.3.1 Germ cell tumours

*Testes:* Testicular neoplasms account for 1% of tumours in males and are classified according to the WHO as follows (Eble et al., 2004):

- Germ cell tumours (94 – 96%).
    i. Seminomatous tumours.
        1. Seminoma, classic type (40-50%).
        2. Spermatocytic seminoma.
    ii. Non seminomatous germ cell tumours (NSGT).
        1. Embryonal carcinoma (5%).
        2. Choriocarcinoma (5%).
        3. Teratoma, mature and immature (5-10%).
        4. Yolk sac tumour.

5. Mixed NSGT.

- Sex cord stromal tumours (4 – 6%).
    i. Leydig cell.
    ii. Sertoli cell.
    iii. Granulosa cell.
    iv. Other.
- Other tumour types (<1%).

Non seminomatous germ cell tumours

Non seminomatous germ cell tumours are thought to recapitulate embryogenesis with their pattern of differentiation being directed toward the formation of one or more of the components of the embryo and/or related structures. The specific direction this differentiation takes will determine the tumour's morphologic appearance thus its name.

1. Embryonal carcinoma: It is composed of primitive carcinoma-like cells with minimal or no signs of differentiation.
2. Mature and immature teratoma (5-10% of all testicular neoplasms): Differentiation is toward structures of the embryo proper, usually a combination of endodermic, mesodermic, and ectodermic tissues.
3. Choriocarcinoma (5% of all testicular neoplasms): This tumour is characterised by the formation of well developed trophoblastic elements. It is associated with a poor prognosis compared with other tumour types.
4. Yolk sac (endodermal sinus) tumour: Differentiation is directed toward the formation of extraembryonic endoderm and mesoderm. It is a tumour of infants and young children.

As differentiation may proceed in a number of directions within the same lesion or there may be only partial differentiation of the components, non seminomatous tumours of the testes are usually mixed tumours with the amount of each component recorded by the pathologist. One of the most common mixtures is embryonal carcinoma and teratoma and this was referred to as a teratocarcinoma.

*Ovary:* Germ cell tumours also occur in the ovary and have a similar classification to testicular tumours as illustrated in Figure 1.8.

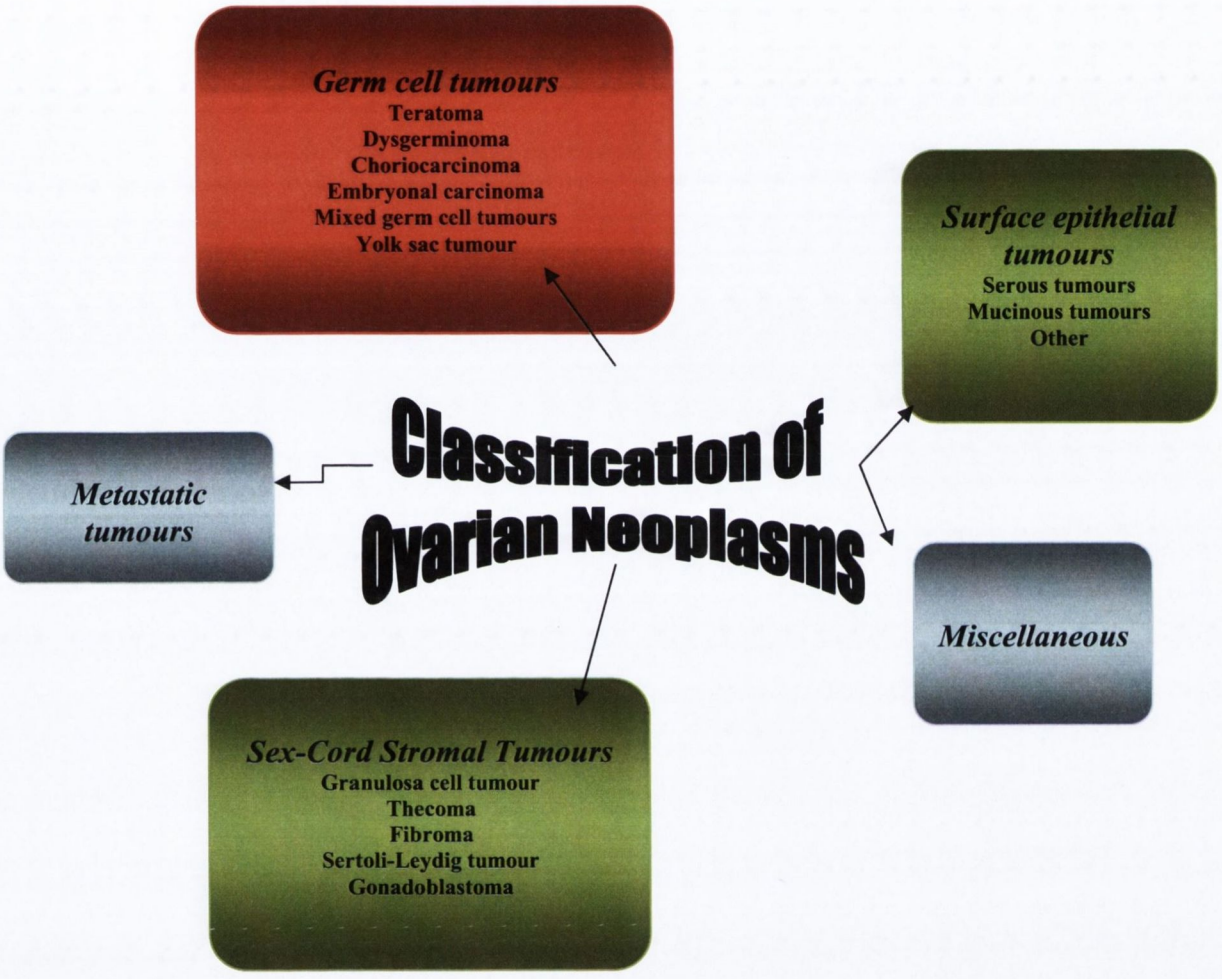



**Figure 1.8: Pathological classification of ovarian tumours.**

### 1.3.2 Embryonal carcinoma

Embryonal carcinoma is a malignant germ cell neoplasm that can occur in the testes or ovary. Within the testis, it occurs most commonly as a component of a mixed non seminomatous germ cell tumour of the testis and is present as a component in more than 80% of mixed germ cell tumours (Bahrami et al., 2007). Pure embryonal carcinoma accounts for only 5% of testicular germ cell neoplasms with an age range at diagnosis of 25 – 35. The presence of

embryonal carcinoma accounting for more than 80% of total tumour volume is reported to be an adverse prognostic factor.

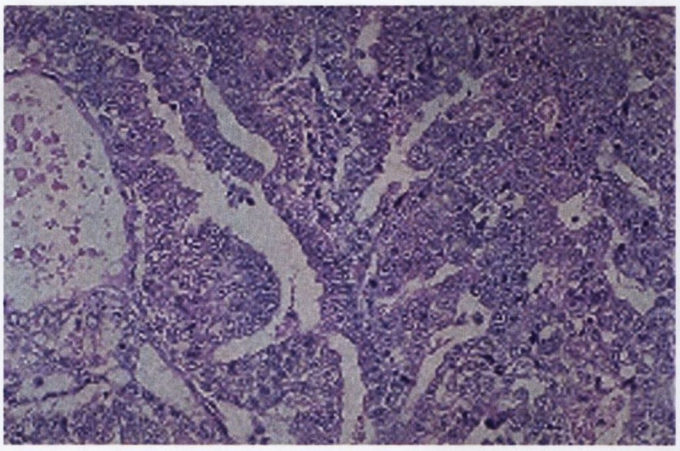


**Figure 1.9: Microscopic appearance of embryonal carcinoma with a tubular and solid growth pattern of large cells with prominent vesicular nuclei and abundant granular cytoplasm. (Taken from Webpath Internet Pathology Laboratory, University of Utah http://library.med.utah.edu/WebPath/webpath.html).**

Embryonal carcinomas in the ovary are rare and occur almost exclusively in children and young adults. They have a similar morphology to embryonal carcinoma of the testis.

### 1.3.3 Teratocarcinomas

Teratocarcinomas are often referred to as the classical stem cell tumours (Andrews, 2002). These tumours contain undifferentiated embryonal carcinoma cells which are the resident cancer stem cells in addition to differentiated cells that represent cells from all three germ layers. Teratocarcinomas present a caricature of embryogenesis and have been subject to investigation for decades (Gudjonsson et al., 2005). The differentiation potential of the teratocarcinoma cells is dramatic; in these tumours one can find the presence of fully differentiated bone, cartilage, brain tissue and cartilage within the tumour mass.

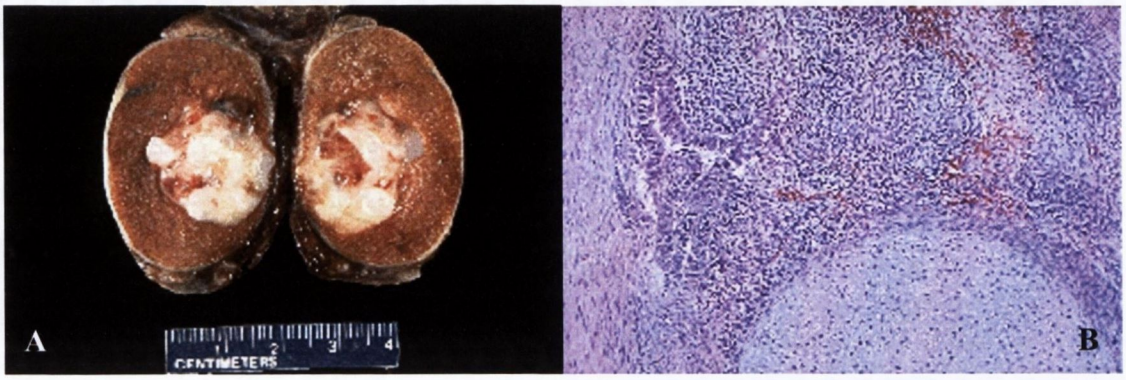



**Figure 1.10: A) Macroscopic appearance of a teratocarcinoma of the testis. The normal testicular parenchyma has a tan brown appearance with a central tumour present, white in colour. B) Microscopic appearance with embryonal carcinoma component and cartilage demonstrating the combination of embryonal carcinoma and teratomatous elements seen in a teratocarcinoma. (Taken from Webpath Internet Pathology Laboratory, University of Utah http://library.med.utah.edu/WebPath/webpath.html)**

The term teratocarcinoma is no longer in widespread clinical use and is a more historical term (Young, 2008). They are now referred to as embryonal carcinomas with teratomatous elements and are as stated previously a common combination within the non seminomatous germ cell tumour category according to the WHO classification of testicular tumours (Eble et al., 2004).

### 1.3.4 Teratomas

Teratomas are germ cell tumours composed of various tissues representing endodermic, mesodermic and ectodermic differentiation. They occur in both the testis and ovary, but are more common in the ovary where they account for the majority of ovarian germ cell neoplasms. The word 'terato' comes from the Greek word 'teras' meaning monster and as a prefix describes an abnormal or monstrous growth. Mature teratomas are composed of entirely mature well – differentiated components as shown in Figure 1.11.

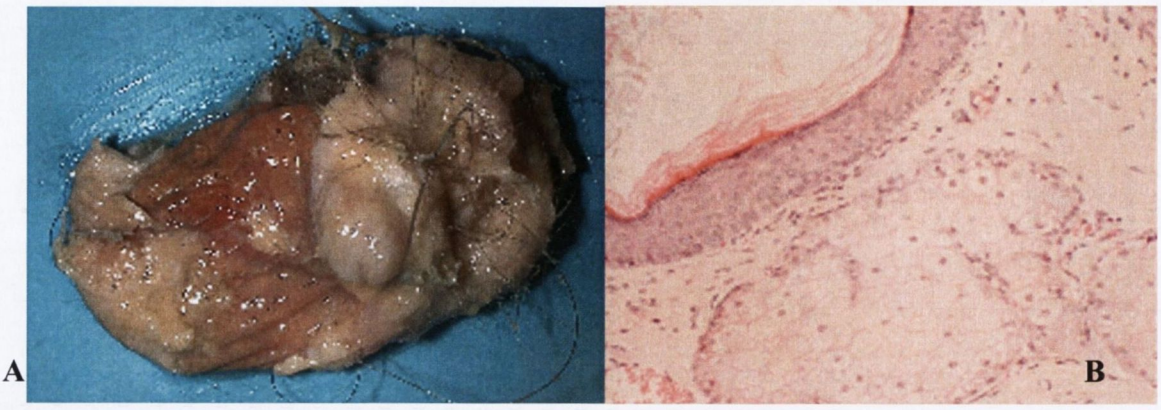



**Figure 1.11: A) Macroscopic appearance of an ovarian teratoma with hair, teeth and skin evident. B) Microscopic appearance showing keratinising squamous epithelium and underlying sebaceous glands reminiscent of fully mature skin.**

Immature teratomas contain embryonic or foetal-like tissues but often accompanied by mature elements. However, the presence of immature elements puts these tumours in the malignant category.

## 1.4 Microarray technology

Microarray technology was introduced in the mid 1990s and offers a tremendous capability for global parallel gene expression monitoring making an ideal starting point for target discovery. The goal of a great majority of current molecular research is the identification of pathways, genes or proteins specific to particular physiological processes and specific disease aetiologies. Great advances have been made in the molecular biology of cancer with the application of microarray technology and the study of differential gene expression profiles of individual tumours. Indeed, it is thought that microarray technology will revolutionise medicine with the provision of treatments and cures for every human disease (Michiels et al., 2007). Enabling personalised cancer treatments is also expected to be possible with the advent of microarray technology and expression profile analysis of tumours (van't Veer et al., 2008). In recent years, it has become apparent that no one gene or protein completely defines the phenotype of a tumour cell; most malignant cells demonstrate disarray of a number of cellular pathways, all of which combine to give the tumour cell its specific characteristics. A summary

of the expression levels within a tumour of all genes involved in these various pathways constitutes the molecular signature of that tumour. Studying molecular signatures rather than specific molecules can give much greater insight into the pathogenesis of tumour development and progression. This has been shown in diffuse large B cell lymphoma (DLBCL), a malignant neoplasm of white blood cells, where the transcription profile of the tumour sub-classified it into two major molecular subtypes, namely the germinal centre B-cell-like (GCB) type of DLBCL and the activated B-cell-like type (ABC) (Leich et al., 2007). These are thought to be two distinct subtypes with different pathogenetic and oncogenic features with the GCB subtype shown to have an improved overall survival compared with the ABC subtype. Breast cancers have also been sub classified using microarray technology into four main groups: a) luminal cell–like (express oestrogen receptor); b) basal cell-like (hormone receptor negative); c) HER-2 positive tumours; and d) normal breast-like group. A comparison of the prognosis between tumour groups was done, the basal-like and HER-2 positive tumours (Sorlie et al., 2001, Reis-Filho et al., 2006).

Functional expression of a gene may be assessed by measuring levels of messenger RNA (mRNA) for that gene. Northern blotting is an effective but labour-intensive and low throughput technique for detecting a single specific mRNA, based on the principle of hybridisation; two nucleic acids strands will hybridise and bind if they are complementary to each other. In Northern blotting, an mRNA sample is run through a gel. A radio-labelled oligonucleotide probe complementary to the desired mRNA target is applied to the gel. If the target mRNA is present, the probe will bind to its location in the gel. The amount of radiation captured on a photographic film is proportional to the amount of target mRNA in the sample (Knudsen, 2004).

DNA microarray analysis is a high-throughput technique analogous to a massively parallel version of Northern blotting. Like Northern blotting, DNA microarrays operate on the basis of probe:RNA hybridisation; however, DNA microarrays utilise tens of thousands of oligonucleotide probes for RNA. They allow rapid comparison of mRNA levels for thousands of genes in multiple biologic samples (Lobenhofer et al., 2001). Thus, instead of running a couple of Northern blots in a day, one can analyse tens of thousands of genes in a dozen or so specimens in a single experiment. This allows simultaneous observation of genes that act in a coordinated fashion, either within the same pathway or in related pathways and genes that

regulate expression of other genes. Thus within the cancer research field, pathways that are consistently dysregulated within a specific tumour type can be identified as well as the genes within each pathway that show altered expression within the tumour allowing insight into the origin of pathway dysregulation and possibly tumour pathogenesis. The end result would be to provide a gene expression profile or gene signature of each different tumour type, a new molecular genetic test in the diagnosis of cancer.

### 1.4.1 Expression microarray technology

Microarrays are created by synthesizing or transferring cDNA (500 – 2000 bases) or oligonucleotide (20 – 70 bases) probes onto a membrane or glass slide. Arrays can have as few as 10 probes up to 2 million depending on the requirements of the researcher. Thousands of probes, each complimentary to a different target sequence, can be attached onto a very small area, often a few $cm^2$. The RNA sample of interest is labelled (typically using fluorescent or chemiluminescent tags) and hybridised to the array. After washing away unhybridised material, the array is excited by a laser and the resultant image scanned. The amount of light emitted from each probe is dependent on the amount of hybridised material; thus the amount of each of the target sequences present in the sample can be quantified from the image generated from the array.

### 1.4.2 Expression microarray studies

Since their inception in 1995, DNA microarrays have revolutionised genome-wide expression profiling of tumours and other tissues (Schena et al., 1995). There has been an exponential rise in the number of microarray papers published since 1995 (Figure 1.12).

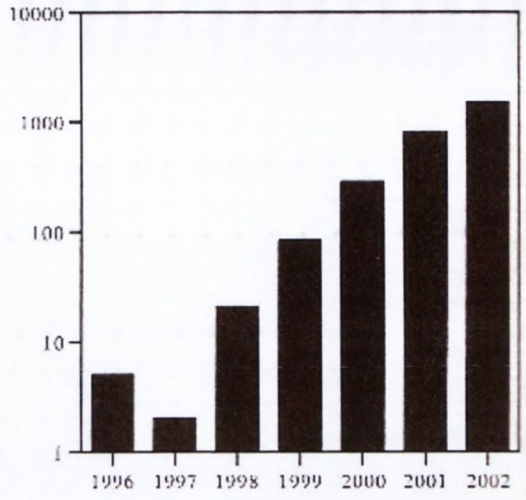



**Figure 1.12: Number of DNA microarray papers published per year (Knudsen, 2004).**

Microarrays have a number of applications in cancer research (Lobenhofer et al., 2001):

1. Determination of differences between malignant cells and benign cells.
2. Refining tumour classification to increase effectiveness of tumour therapies.
3. Identification of genes implicated in tumour formation or progression.
4. Development of personalized cancer treatments depending on individual tumour gene expression profiles.

### 1.4.3 Applied Biosystems Expression Array System

The Applied Biosystems expression array system is an expression profiling system that combines highly sensitive chemiluminescence detection chemistries with a fully curated transcript data set with comprehensive coverage of the total mouse and human genomes. The Applied Biosystems array platform employs single 60-mer oligonucleotides and uses a single-colour chemiluminescence detection technology. The Applied Biosystems Mouse Genome Survey Microarray contains probes representing a complete, annotated, and fully curated set of approximately 32,000 mouse genes from the public and Celera databases. The Applied Biosystems Human Genome Survey Microarray provides 31,700 probes for interrogation of 27,868 genes. The system includes a complete searchable database containing Celera

Discovery System™ and public database annotations for transcript sequences represented on the array. Each array offers extensive data reproducibility and quality control measured by chemiluminescent and fluorescent internal controls. The system integrates seamlessly with TaqMan® Gene Expression Assays and TaqMan® Low Density Custom Arrays for individual target validation.

### 1.5 Objectives of this study

Embryonic stem cells have been a subject of much interest given their capacity for self-renewal and for differentiation into desired cell types, particularly in the field of regenerative medicine where they provide hope for new cell replacement therapies. Embryonic stem cells have another significant property – their similarities with cancer cells and in particular cancer stem cells. The study and analysis of their similarities and differences offers a vast minefield of targets that may be of potential use in future cancer therapies. Ultimately, the identification of a target and or treatment that would eliminate cancer stem cells would constitute a major advance in the treatment of cancer and prevention of recurrence or metastasis of tumours. Much research is being currently undertaken to investigate the properties of stem cells, cancer stem cells and their role in carcinogenesis.

This study sets out to investigate the similarities and differences between embryonic stem cells and their malignant equivalent, embryonal carcinoma cells (cancer stem cells) based on the model of teratoma tumourigenesis in mouse and human. This study also sets out to obtain transcriptome profiles of the fundamental properties of stem cells, i.e. self-renewal and differentiation, with the ultimate aim of providing a transcriptome profile of the cancer stem cell and highlighting potential targets for future cancer therapies. As disordered differentiation and dysregulation of self-renewal are considered to be the basis of carcinogenesis, this study focuses on an early differentiation time point (day 3), unlike other studies involving much later time points, in an attempt to isolate the earliest neoplastic genetic event.

To achieve these aims, a cell culture model and differentiation time course was designed and extensively tested using mES and mouse EC cells (both pluripotent and nullipotent forms).

Cells were grown for 3 days in their undifferentiated and differentiated states, a significantly earlier time point than most studies currently in the literature, and RNA was harvested. Subsequent expression array analysis was performed using Applied Biosystems technology, a technology platform not widely published on in the area of stem cell biology. An equivalent human model was also established using nulli- and pluripotent human EC cell lines, again using the early differentiation time point of 3 days.

# CHAPTER 2 MATERIALS AND METHODS

## 2.1 Mouse Cell lines

### 2.1.1 Mouse nullipotent teratocarcinoma cells, Nulli-SCC1, ATCC number: CRL 1566

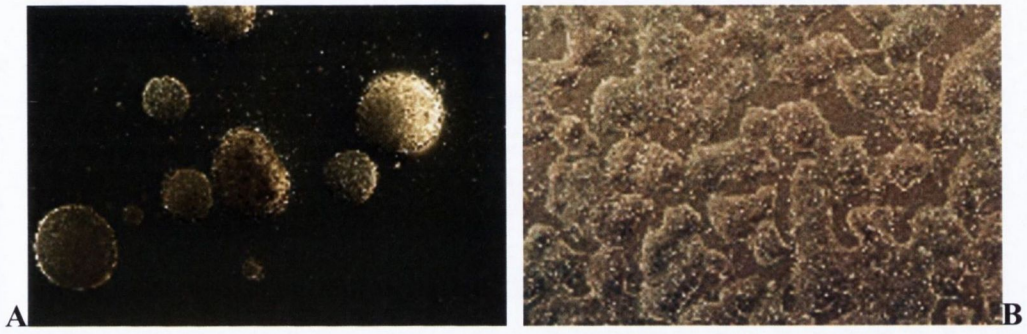



**Figure 2.1: A) Nulli-SCC1, Phase contrast. B) Nulli-SCC1, Day 3, T75 flask, 5x, phase contrast.**

This cell line from ATCC (American Type Culture Collection (www.atcc.org)) was deposited by Gail Martin and derived from a murine testicular teratocarcinoma. These cells do not differentiate, but can be made to form aggregates and can serve as control for pluripotent teratocarcinoma cell lines such as SCC-PSA1 (see below). These cells grow best on gelatine coated plates and form flattened colonies. Medium required for growth is Dulbecco's modified Eagle's medium (DMEM) made up as shown in Table 2.1. All cells were grown in a 37°C incubator with 5% $CO_2$ and appropriate humidity.

**Table 2.1: Culture medium components (All Gibco cell culture components obtained from Invitrogen Corporation, Carlsbad, CA, USA).**

| Component | Product details | Volume (ml) |
|---|---|---|
| DMEM | Gibco 21969-035 | 430 |
| Foetal bovine serum (10%) | Gibco 16000-044 | 50 |
| L-Glutamine (4mM) | Gibco 25030-024 | 10 |
| Penicillin/Streptomycin | Gibco 15140-122 | 10 |
| *Total* | | *500* |

### 2.1.2 Mouse pluripotent teratocarcinoma cells, SCC-PSA1, ATCC number CRL 1535

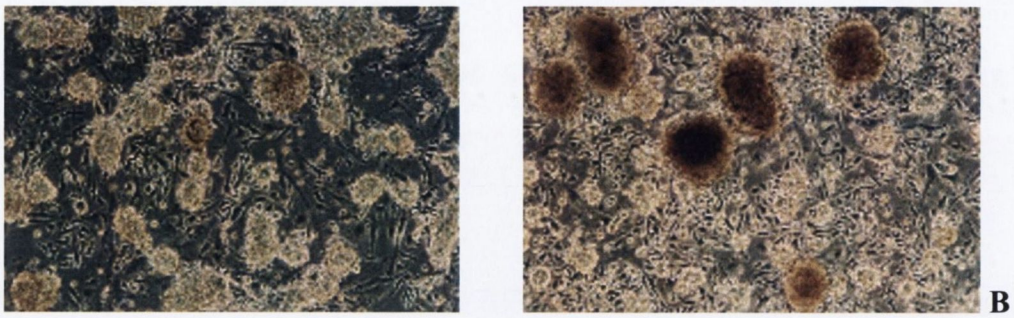



**Figure 2.2: A) SCC-PSA1, D0 (T75 on feeders), 10x. B) SCC-PSA1, Day 3, differentiated (tissue culture dish), 10x.**

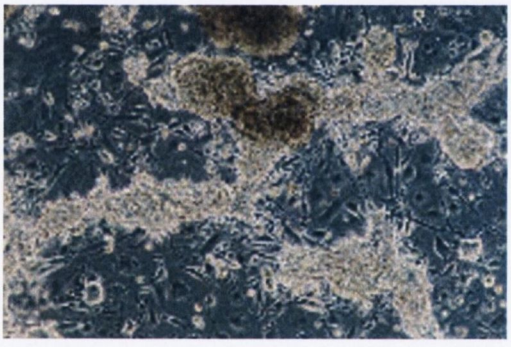


**Figure 2.3: SCC-PSA1, Day 3, undifferentiated (T75), 10x.**

This cell line from ATCC was deposited by Gail Martin and derived from a murine testicular pluripotent teratocarcinoma (isolated from secondary cultures of the OT/5568 transplantable tumour). Cells were maintained in an undifferentiated state by frequent subculture and the use of a fibroblast feeder cell layer (see section 2.1.4 below). In the absence of a feeder layer, the cells begin to differentiate. Medium required for growth is Dulbecco's modified Eagle's medium made up as shown in Table 2.1. All cells were grown in a 37°C incubator with 5% $CO_2$ and appropriate humidity.

### 2.1.3 Mouse embryonic stem cells, ES-E14TG2a, ATCC number CRL-1821.

This cell line from ATCC was deposited by T. Doetschman and is a derivative of one of several embryonal stem cell lines developed by M. Hooper in 1987. Cells are deficient in HgPRT and are resistant to 0.06 mM 6-thioguanine.

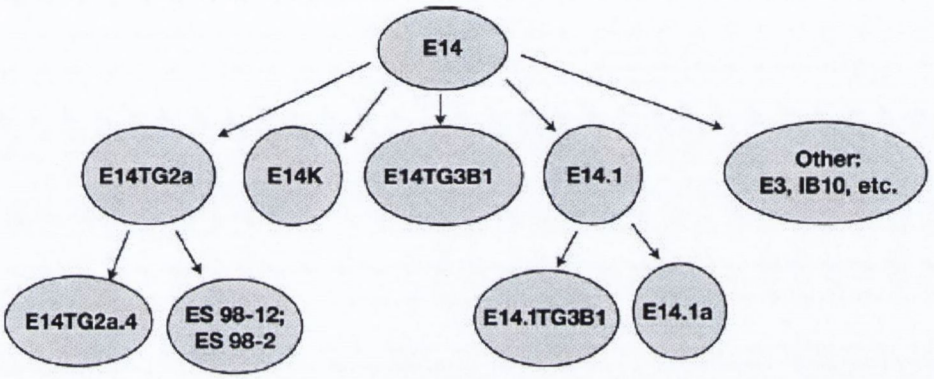



**Figure 2.4: Sub clones and cell lines derived from E14, a key mouse embryonic stem cell lines used widely in the literature (Downing et al., 2004).**

Cells remain undifferentiated when cultured on feeder layers. In the absence of a feeder layer, cells spontaneously differentiate and form embryonal structures. Medium required for growth is Dulbecco's modified Eagle's medium made up as shown in Table 2.2. All cells were grown in a 37°C incubator with 5% $CO_2$ and appropriate humidity.

**Table 2.2: Culture medium components for mouse embryonic stem cells (all Gibco products obtained from Invitrogen, Carlsbad, CA, USA while 2-Mercaptoethanol was obtained from Sigma-Alrich Ireland Ltd, Dublin, Ireland).**

| Component | Product details | Volume |
|---|---|---|
| DMEM | Gibco 21969-035 | 430ml |
| Foetal bovine serum (10%) | Gibco 16000-044 | 50ml |
| L-Glutamine (4mM) | Gibco 25030-024 | 10ml |
| Penicillin/Streptomycin | Gibco 15140-122 | 10ml |
| 2-Mercaptoethanol (0.1mM) | Sigma M6250 | 3.5µl |
| *Total* | | *500ml* |

### 2.1.4 Irradiated mouse embryonic fibroblasts, Source culture STO (CRL-1503), ATCC number 56-X

This cell line from the ATCC was deposited by Gail Martin. Its source culture is STO (CRL-1503), a cell line of mouse embryonic fibroblasts derived by A. Bernstein, Canada. This cell line is routinely used to prepare feeder layers by irradiation as in the current case or mitomycin C treatment, therefore they will not divide or replicate. Medium required for growth was Dulbecco's modified Eagle's medium made up as shown in Table 2.1. All cells were grown in a 37°C incubator with 5% $CO_2$ and appropriate humidity.

### 2.1.5 Preparation of gelatine coated flasks

Tissue culture flasks (T25 and T75, Sarstedt, Nümbrecht, Germany) for use with all 4 mouse cell lines were required to have a gelatine coating prior to use. Tissue culture dishes (9cm) and bacteriological dishes used with these cell lines did not require gelatine coating. A 1% solution of gelatine was prepared in advance made in 50ml batches (50ml distilled water, 0.5g gelatine powder). The gelatine was melted into solution for approximately 20 seconds in the microwave on the high setting. The gelatine solution was then autoclaved for 10 minutes at 121°C. It was then ready for use. It was stored at 4°C.

Gelatine was added to flasks 30 minutes prior to use with 1ml per T25 and 3ml per T75 flask and left to stand. Used gelatine was discarded after use.

### 2.1.6 Preparing feeder layered flasks

Vials of irradiated mouse fibroblasts were purchased from ATCC to coat flasks prior to seeding with either mouse pluripotent teratocarcinoma cell line, SCC-PSA1 or mouse embryonic stem cell line, ES-E14TG2a. One vial was used per T75 or two T25 flasks. Fibroblasts were seeded at least 2 days prior to seeding of appropriate secondary cell line and should have reached at least 80% confluency by this stage. Standard thawing procedures were

used prior to seeding of these cells and the standard DMEM formulation was used as medium. Flasks used were gelatine coated.

### 2.1.7 Seeding of cell lines

All cell lines were seeded in a similar fashion. One vial of cells was removed from storage in liquid nitrogen and thawed in a 37°C water bath for 2 minutes. Cells were then transferred to a 15ml tube containing 8ml of pre-warmed (37°C) medium. Cells were centrifuged at 150g for 8 minutes and resuspended in 1ml of fresh pre-warmed medium. These cells were seeded into 2 T25 flasks. Medium was changed on day 2 post seeding. On day 3, cells were passaged with cells from each T25 were transferred into a T75 flask.

### 2.1.8 Subculture/passage of cell lines

Every 3 days, cells were subcultured. Spent medium was removed. Cells were rinsed with 1X phosphate buffered saline (PBS) (Gibco) prior to addition of trypsin 0.25%/EDTA0.03% (Gibco Catalogue Number 25200-072), 1ml per T25 and 3ml per T75 flask. All reagents were pre-warmed in a 37°C water bath. Following addition of trypsin, cells were incubated in the tissue culture incubator for 4 minutes. The trypsin was subsequently neutralised with an equivalent volume of medium. Cells were transferred to a 15ml tube and centrifuged at 150g for 8 minutes. Cells were resuspended in 1ml of fresh medium. Cells were split 1 in 4 apart from the mouse embryonic stem cells, ES-E14TG2a, which were split 1 in 6.

### 2.1.9 Differentiation protocol

The differentiation protocol followed was that recommended by the ATCC when growing the SCC-PSA1 cell line. The first step was initiation of embryoid body formation where on day 3, cells were harvested and seeded at $1x10^7$ cells/9cm tissue culture dish without feeders. Medium was changed on day 2 after plating. Further differentiation (embryoid body development in suspension) was performed by transferring cells at day 3 to a bacteriological

Petri dish with a harvest from one 9cm tissue culture dish transferred to 10ml of medium in a 9cm bacteriological dish. Medium was changed every second day, with a fresh bacteriological dish required every 3 days. To allow subsequent differentiation, embryoid bodies having been grown in suspension for at least 5 days were transferred to 9cm tissue culture dishes without feeders or gelatine. Medium was changed every second day. When harvesting or transferring cells from these dishes, fresh medium was gently pipetted over the clumps of cells to loosen them. Medium containing clumps was collected and transferred to a conical tube. Second and third batches of fresh medium to loosen clumps were used as required. All washes were pooled and the clumps were allowed to settle. Spent medium was removed and cells were either harvested and frozen at -80°C or resuspended in fresh medium. This protocol was used with all mouse cell lines.

### 2.1.10 Protocol to maintain undifferentiated cultures

Cells were passaged every three days or frequently enough to prevent them from forming large piled up colonies as opposed to flattened colonies. Cells were seeded 1 in 4 or 1 in 6 in gelatinised flasks with a feeder layer. Medium was changed every second day after plating with harvesting on the third day of incubation. This protocol was used with all mouse cell lines. However, the Nulli-SCC1 cell line was not grown on feeder layers at any stage, but were treated the same in every other respect.

### 2.1.11 Cryopreservation of cells

Aliquots of cells to be frozen and stored in liquid nitrogen were frozen in 1ml of freeze medium (95% DMEM and 5% Dimethyl sulphoxide (DMSO), Gibco Catalogue number 11101-011), following resuspension in fresh medium after harvesting.

Cells pellets to be used for RNA extraction were transferred to 1.5ml eppendorf tubes and centrifuged at 150g x 5 minutes. Any remaining spent medium was removed prior to labelling and storage of cell pellets at -80°C.

### 2.1.12 Additional requirements for mouse embryonic stem cells, ES-E14TG2a

Embryonic stem cells can be difficult to grow so some additional measures may be used to encourage growth. The following were used:

- Medium was changed every day rather than every second day as with other cell lines.
- Medium was changed 3-4 hours prior to passage.
- Following trypsinisation of cells, cells were pipetted up and down repeatedly to ensure a single cell suspension prior to neutralisation and resuspension.

### 2.1.13 Cell counting using a haemocytometer

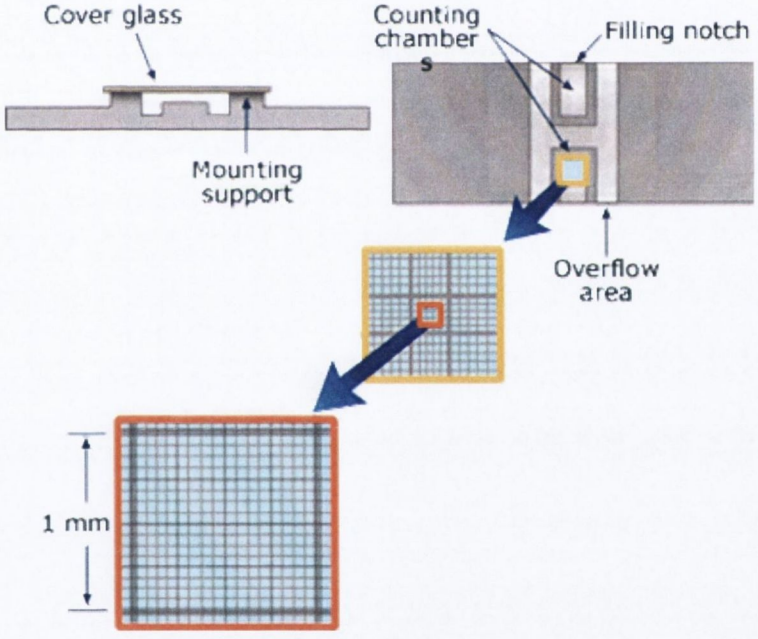



**Figure 2.5: Illustration of a Haemocytometer (Taken from http://toolboxes.flexiblelearning.net.au/demosites/series4/412/laboratory/studynotes/SNHaemo.htm).**

A haemocytometer was used to obtain an accurate count of cells and to determine the suspension counts for all cell strains. It consists of a thickened glass slide into which a small chamber has been cut to allow for the introduction of cells to be counted. The floor of the chamber is divided (etched) into nine sections; usually only the four corner sections are used

in cell counting. With a coverslip in place, each square of the haemocytometer represents a total volume of $0.1mm^3$ or $10^{-4}cm^3$. Since $1cm^3$ is approximately equivalent to 1ml, the cell concentration per ml (and the total number of cells) can be determined.

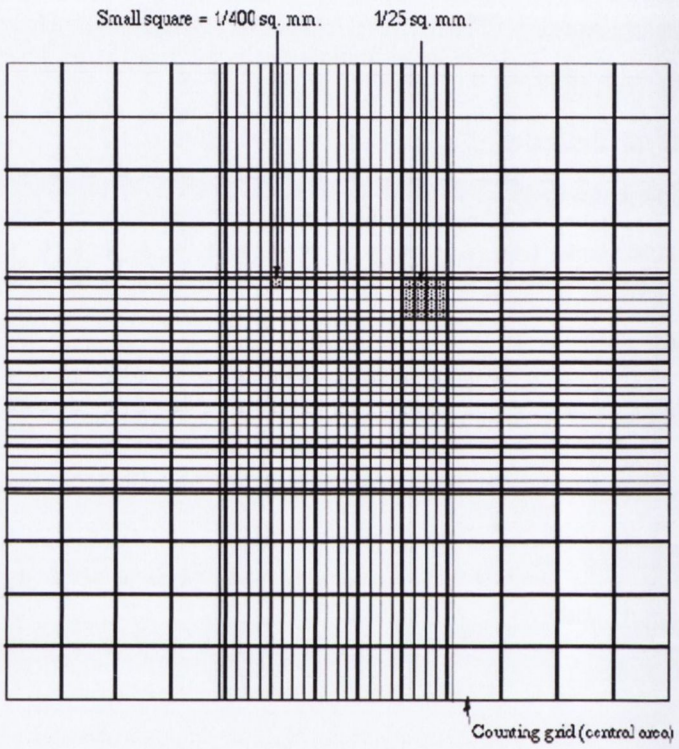



**Figure 2.6: View of haemocytometer when visualised down the microscope (Taken from Wikipedia, http://www.wikipedia.org/).**

- The haemocytometer was prepared by cleaning all surfaces and making sure it was completely dry using non-linting tissue.
- The coverslip was centred on the haemocytometer.
- 20µl of the cell suspension to be counted was removed and added to 20µl of trypan blue (Gibco catalogue number 15250-061) and mixed allowing it to settle for 1-2 minutes.
- 9µl of the trypan blue/cell suspension was pipetted into one of the two counting chambers, filling them slowly and steadily. It was important not to overfill or under fill the chambers.
- The cell suspension was allowed to settle for at least 10 seconds.

- The cells were counted in the four corner squares, counting the cells touching the top or left borders but not those touching the bottom or right borders.
- The cell count was calculated using the equation: cells/ml = (n) x dilution factor (2 in all my experiments) x $10^4$ where n = the average cell count per square of the four corner squares counted.
- The total number of cells was determined in the total suspension volume by multiplying the volume of the cell suspension by the 'cells/ml' value already calculated.

### 2.1.14 Heat inactivation of Foetal Bovine Serum (FBS)

Foetal bovine serum (stored at -20°C) was required to be heat inactivated before use in cell medium. It was thawed overnight at 4°C. A water bath was preheated to slightly higher than 56°C to allow for cooling when the serum was added. The serum was placed in the water bath to ¼ inch above the serum line in the bottle. The serum remains in the water bath for 30 minutes. Following this, it was cooled rapidly. It was divided into 25ml aliquots and stored at -20°C.

### 2.1.15 Use of Leukaemia inhibitory factor (LIF)

LIF was purchased from Sigma, catalogue number L5158, in vials (1ml vials containing 106 Units/ml) that can be stored between 2-8°C for up to 1 year. One vial or 106 Units/ml was required per 500ml of cell culture medium and was added when required when making up media in 500ml batches. LIF, a multifunctional cytokine belonging to the interleukin-6 family, has been used as a culture medium additive to maintain mES cells in their undifferentiated state, thus removing the requirement for fibroblast feeder layers.

### 2.1.16 Mycoplasma testing of cell culture

Regular testing for mycoplasma contamination of all cell lines was performed using the mycoplasma PCR Elisa assay from Roche Diagnostics GmbH, Mannheim, Germany (Catalogue number 1 663 925). This assay detects all mycoplasma. DNA from other bacteria, yeast and eukaryotic cells is not detected. The sensitivity is at least 1 x $10^3$ colony-forming units (cfu) of mycoplasma per millilitre of culture medium. The assay is based on the amplification of a mycoplasma-specific DNA sequence by PCR and the subsequent detection of the amplicon by ELISA. Mycoplasma contained in the sample are enriched by centrifugation of the cell-free supernatant, then lysed. A conserved and mycoplasma -specific region of the DNA is amplified by PCR in the presence of DIG-labelled dUTP. The DIG-labelled amplicon is denatured by NaOH treatment, hybridised to a biotin-labelled capture probe, and immobilised on a streptavidin-coated microplate. The immobilised amplicon is detected by means of an anti-digoxigenin, coupled to horseradish peroxidase, antibody (anti-DIG-POD), and the sensitive peroxidase substrate TMB.

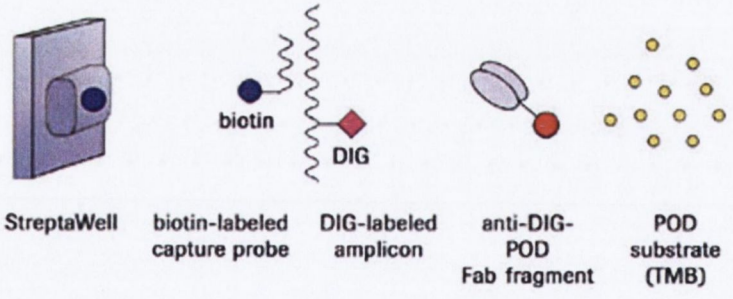



**Figure 2.7: Mycoplasma assay principle (Taken from Mycoplasma PCR ELISA instruction manual from Roche, February 2006).**

At least two negative controls and one positive control were used in each experiment.

- 1ml culture supernatant was transferred into a centrifugation tube.
- The sample was centrifuged at 200g x 10minutes at 15-25°C to pellet cells and any cellular debris present.
- The supernatant was transferred into a fresh sterile tube and centrifuged at 13,000g x 10minutes at 2-8°C to sediment any mycoplasma.

- The supernatant was decanted avoiding loss of pellet if any.
- 10µl of double distilled sterile water and 10µl of lysis reagent (bottle 1) were added to the pellet and the material dissolved.
- For the positive control, 10µl of positive control DNA (vial 12) along with 10µl of lysis reagent were added to a tube.
- For the two negative controls, 10µl of double distilled water and 10µl of lysis reagent were used.
- Samples incubated at 37°C for 1 hour.
- 30µl of neutralisation agent (bottle 2) added to each sample.
- 25µl of PCR ready-to-go mix (vial 3) and 15µl of sterile water were added to a tube and 10µl of the sample or control is added.
- Samples were transferred to a thermal cycler and the following thermal cycle performed:

**Table 2.3: PCR conditions for ELISA.**

| Cycle Number | Conditions |
|---|---|
| **Cycle 1** | 5 min, 95°C |
| **Cycle 2-40** | 30s, 94°C; 30s, 62°C; 1min, 72°C |
| **Cycle 41** | 10min, 72°C |
| **Hold** | 8°C |

- 40µl of denaturation reagent (bottle 4) was added to a reaction tube.
- 10µl of the amplification product was added and incubated at 15-25°C for 10 minutes.
- 450µl of hybridisation agent was added and mixed thoroughly.
- 200µl of the mixture was added to a well in the microplate (performed in duplicate).
- Wells were covered by a self-adhesive cover foil and incubated at 37°C on a shaker (300rpm) for 3 hours.
- The hybridisation mixture was removed by flicking and washing three times with 250µl of washing buffer (1x) per well. This was also removed by flicking.
- 200µl of anti-DIG-POD working dilution was added per well.
- The wells were covered by a self-adhesive cover foil and incubated at 15-25°C on a shaker (300rpm) for 30 minutes.

- The anti-DIG-POD was removed by flicking and washed five times with 250µl of washing buffer (1x) per well. This was also removed by flicking.
- 100µl of TMB substrate solution (bottle 10), prewarmed to 15-25°C, added per well.
- The wells were covered with a self-adhesive cover foil and incubated at 15-25°C on a shaker (300rpm) for 20 minutes.
- 100µl of stop reagent (bottle 11) was added per well and the absorbance was measured using a microplate (ELISA) reader at 450nm (reference wavelength approximately 690nm) within 1 hour after addition of the Stop reagent.
- Interpretation of results:
    - The absorbance of the negative control should be lower than 0.25 $A_{450nm} - A_{690nm}$ – units.
    - The absorbance of the positive control should be higher than 1.2 $A_{450nm} - A_{690nm}$ – units.
    - The mean of the absorbance readings of the negative controls was subtracted from those of the samples. The samples were regarded as contaminated if the difference in absorbance is higher than 0.2 $A_{450nm} - A_{690nm}$ – units.

## 2.2: Human cell lines

### 2.2.1 Human teratocarcinoma cells, pluripotent, Ntera-2 clone D1 – NT2/D1

Designation: Ntera-2 clone D1 – NT2/D1

This cell line was received as a gift from Professor Peter Andrews, Department of Biochemical Science, University of Sheffield, UK. He has also deposited this cell line at the ATCC®, ATCC® number CRL-1973™ where the rest of the cell lines used were purchased. This cell line is derived from a human testicular malignant embryonal carcinoma. A primary characteristic of this cell line is its ability to maintain its stemness for unlimited periods *in vitro*. This clone differentiates along neuroectodermal lineages after exposure to retinoic acid (RA) or hexamethylene bisacetamide (HMBA) (Andrews, 1998). The RA induced differentiation is characterized by glycolipid changes, appearance of neurons, and induction of

homeobox (HOX) gene clusters. The cells exhibit high expression of N-myc oncogene activity.

**Table 2.4: Media requirements for human pluripotent teratocarcinoma cells, NT2/D1.**

| Component | Product details | Volume |
|---|---|---|
| DMEM (containing no Na pyruvate, high glucose) | Gibco 12491-015 | 430ml |
| Foetal bovine serum (10%) | Gibco 16000-044 | 50ml |
| L-Glutamine (4mM) | Gibco 25030-024 | 10ml |
| Penicillin/Streptomycin | Gibco 15140-122 | 10ml |
| *Total* | | *500ml* |

Medium required for growth is Dulbecco's modified Eagle's medium made up as shown in Table 2.4. All cells were grown in a 37°C incubator with 5% $CO_2$ and appropriate humidity. Medium when made up was stored at 4°C and used within 3 days. Flasks do not need to be gelatinised as they were for the murine cell lines.

#### *2.2.1.1 Subculture/Passage of cell lines:*

NT2/D1 cells were split by scraping to maintain their phenotype using glass beads or cell scrapers. It is possible to use either method, the important thing is that the cells like to be in small clumps. NT2/D1 were passaged when quite heavily confluent, approximately every 3-4 days. A 1:3, 1:4 split is ideal, about $5x10^6$ or more cells per T75. Medium was removed, leaving about 2ml in the flask. A cell scraper was then used to remove the adherent cells. The cells came off in clumps. The cell suspension was then pipetted up and down to break the clumps up slightly, before dispensing into new T75 flasks with 25ml of medium. All reagents were pre-warmed in a 37°C water bath.

### *2.2.1.2 Differentiation of NT2/D1*

Medium was DMEM/FBS (as per table) with Retinoic acid added at a concentration of $10^{-5}$M (see section 2.1.3). Cells were fed every 7 days with this new medium.

**Table 2.5: Medium used for differentiation of NT2/D1. (Gibco products obtained from Invitrogen, Carlsbad, CA, USA, Sigmas products obtained from Sigma-Alrich Ireland Ltd, Dublin, Ireland.)**

| Component | Product details | Volume |
|---|---|---|
| DMEM (containing no Na pyruvate, high glucose) | Gibco 12491-015 | 430ml |
| Foetal bovine serum (10%) | Gibco 16000-044 | 50ml |
| L-Glutamine (4mM) | Gibco 25030-024 | 10ml |
| Penicillin/Streptomycin | Gibco 15140-122 | 10ml |
| Retinoic acid, $1x10^{-5}$M (see section 2.2.4) | Sigma | 50µl |
| Total | | 500ml |

### *2.2.1.3 Making up retinoic acid for differentiation protocol*

Requirements:

- All-trans Retinoic acid, Sigma, R2625.
- DMSO, Sigma, D2438.

The all-trans retinoic acid can be bought in 50mg, 100mg, 500mg, 1g and 5g quantities as a yellow powder. This powder was stored in liquid nitrogen and can be stored long term. The DMSO used was pure DMSO and comes in 10ml or 50ml quantities. It was stored at room temperature. Once opened, the remainder must be discarded.

Retinoic acid for use in these experiments was made up as follows:

- 100mg all-trans Retinoic acid was dissolved in 33ml DMSO, 0.01M.
- It was stored in 1ml aliquots at -80°C.

- Aliquots in use were stored at -20°C.
- When needed, aliquots were thawed at room temperature away from light.
- Thawing and refreezing does not appear to affect the retinoic acid.
- A 1:1000 dilution was required in the medium (i.e. $10^{-5}$M), i.e. 5µl per 50ml of media.

#### *2.2.1.4 Harvesting of NT2/D1*

NT2/D1 cells were harvested using Trypsin/EDTA. Spent medium was removed. Cells were rinsed with 1X PBS (phosphate buffered saline, Gibco) prior to addition of trypsin 0.25%/EDTA0.03% (Gibco Cat No 25200-072), 1ml per T25 and 3ml per T75 flask. All reagents were prewarmed in a 37°C water bath. Following addition of trypsin, cells were incubated in the tissue culture incubator for 4 minutes. The flask was tapped and the cells resuspended in 10ml medium. A cell count was performed. Cells were reseeded at 1- $2x10^6$ cells per T75 in 12ml of medium.

#### *2.2.1.5 Cryopreservation of NT2-D1 cell line*

Aliquots of cells to be frozen and stored in liquid nitrogen were frozen in 1ml of freeze medium (95% DMEM and 5% DMSO, Gibco Cat no 11101-011), following resuspension in fresh medium after harvesting.

Cells pellets to be used for RNA extraction were transferred to 1.5ml eppendorf tubes and centrifuged at 150g x 5 minutes. Any remaining spent medium was removed prior to labelling and storage of cell pellets at -80°C.

### 2.2.2 Human teratocarcinoma cell line, nullipotent, 2102Ep Cells

Designation: 2102Ep

This cell line was received as a gift from Professor Peter Andrews, Department of Biochemical Science, University of Sheffield, UK.

**Table 2.6: Medium used for culture of 2102Ep cells. (Gibco products obtained from Invitrogen, Carlsbad, CA, USA.)**

| Component | Product details | Volume |
|---|---|---|
| DMEM (containing no Na pyruvate, high glucose) | Gibco 12491-015 | 430ml |
| Foetal bovine serum (10%) | Gibco 16000-044 | 50ml |
| L-Glutamine (4mM) | Gibco 25030-024 | 10ml |
| Penicillin/Streptomycin | Gibco 15140-122 | 10ml |
| *Total* | | *500ml* |

### *2.2.2.1 Subculture of 2102Ep*

This cell line was passaged every 3 - 4 days until flasks were fully confluent. Trypsinisation was the method used for passaging and harvesting of cells, unlike the NT2/D1 cell line which uses cell scrapers alone.

- Trypsinised with Trypsin/EDTA, 1-2ml per T75.
- Incubated for 3-4 minutes in the incubator at 37°C.
- On removal, the flask was knocked to dislodge the cells.
- Cells were resuspended in 10ml medium.
- Cells titrated up and down with pipette.
- Dispensed into new flasks, 1:3 or 1:4 split—at least $5x10^6$cells per T75 flask.

### *2.2.2.2 Cryopreservation of 2102Ep cell line*

Aliquots of cells to be frozen and stored in liquid nitrogen were frozen in 1ml of freeze medium (95% DMEM and 5% DMSO, Gibco Cat no 11101-011), following resuspension in fresh medium after harvesting.

Cells pellets to be used for RNA extraction were transferred to 1.5ml eppendorf tubes and centrifuged at 150g x 5 minutes. Any remaining spent medium was removed prior to labelling and storage of cell pellets at -80°C.

## 2.3: Total RNA extraction methods

### 2.3.1 Extraction of total RNA from cell pellets

Extraction of total RNA from cell pellets was performed using the Qiagen RNeasy® Mini kit, Catalogue number 74104 (Qiagen Ltd, West Sussex, UK). The protocol for isolation of total RNA from animal cells was used. The cell pellets were loosened thoroughly by flicking the tube. The number of pelleted cells always ranged between 5 x $10^6$ and 1 x $10^7$ so as to disrupt the cells, 600µl of Buffer RLT was added to each cell pellet. β-mercaptoethanol (β-ME) was added to Buffer RLT before use, 10µl β-ME per 1ml Buffer RLT. Cells were pipetted to mix and thoroughly disrupt the cells. To homogenise the sample, the lysate was applied directly onto a QIAshredder spin column which had been placed in a 2ml collection tube. The samples were then centrifuged for 2 minutes at maximum speed. 600µl of 70% ethanol was added to the homogenised lysate and mixed well by pipetting only. Aliquots of 700µl of sample were loaded successively on the RNeasy® mini column placed in a new 2ml collection tube. They were centrifuged for 15s at > 8000g. The flow-through was discarded. In the next step, 700µl of Buffer RW1 was added to the RNeasy® column and centrifuged for 15s at > 8000g to wash the column. Both the flow-through and collection tube were discarded. After transferring the RNeasy® column into a new 2ml collection tube, 500µl of Buffer RPE was added onto the RNeasy® column and centrifuged for 15s at > 8000g to wash the column. A further 500µl of Buffer RPE was added to the column and centrifuged for 2 minutes at > 8000g to dry the RNeasy® silica-gel membrane. The RNeasy® column was placed in a new 2ml collection tube and centrifuged at full speed for 1 minute. To elute, the RNeasy® column was transferred to a new 1.55ml collection tube. A volume of 50µl of RNase-free water was pipetted directly onto the RNeasy® silica-gel membrane and centrifuged for 1 minute at > 8000g. The column was then discarded and the samples were labelled and placed on ice ready to proceed with DNase treatment. Alternatively, the RNA was stored at -80C for future use.

RNeasy® technology simplifies total RNA isolation by combining the stringency of guanidine-isothiocyanate lysis with the speed and purity of silica-gel–membrane purification. There is no need for centrifugation through cesium chloride (CsCl) cushions or precipitation with lithium chloride (LiCl) or alcohol. RNeasy® Kits provide the ability to produce high quality RNA with minimum copurification of DNA. For certain RNA applications that are sensitive to very small amounts of DNA, the residual amounts of DNA remaining can be removed using the DNase treatment during the RNeasy® procedure or following it using alternative DNase treatments.

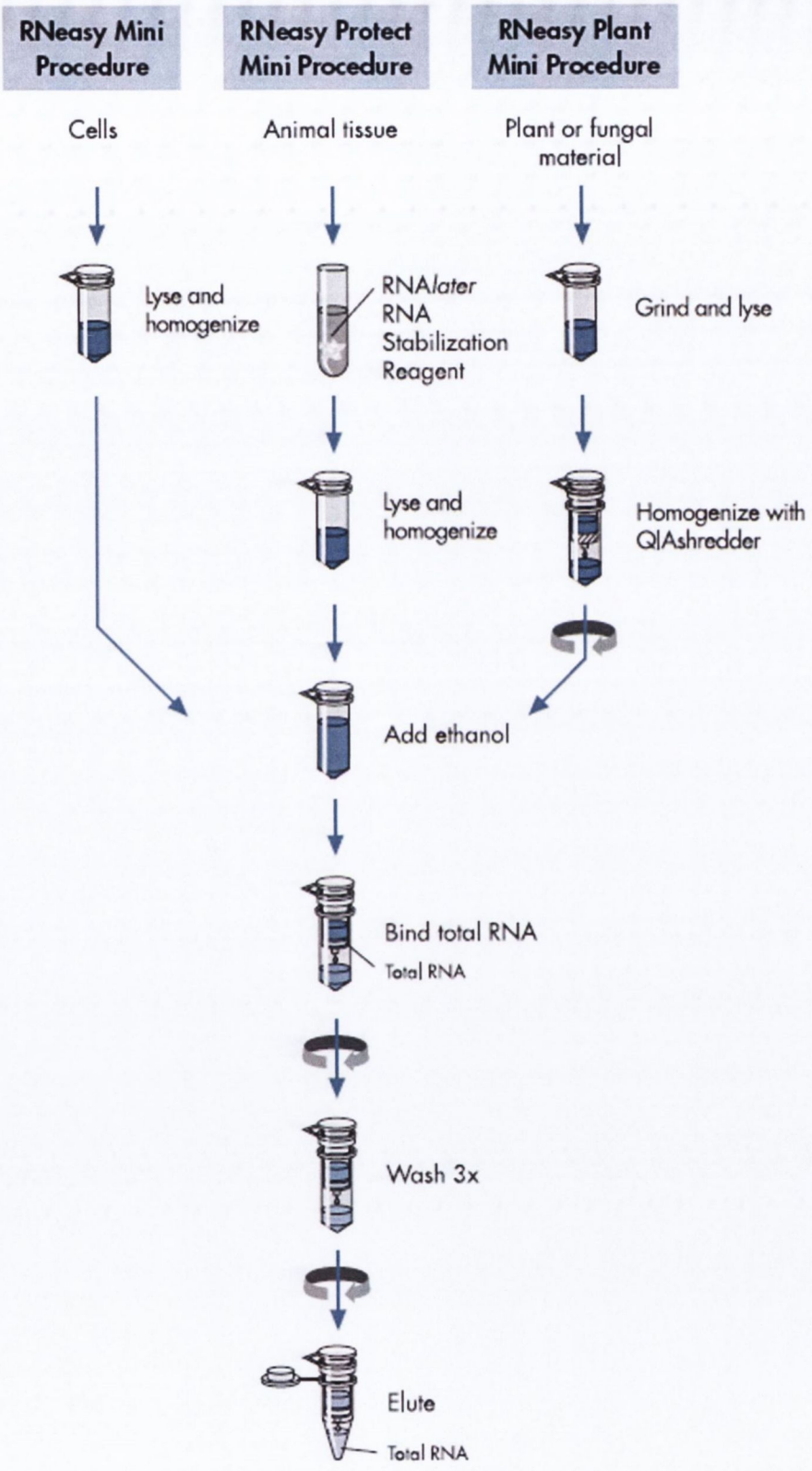



**Figure 2.8: RNA extraction procedure using RNeasy® minikit (Taken from the RNeasy® Mini Handbook from Qiagen, 4th Edition, April 2006).**

### 2.3.2 DNase treatment of Total RNA samples

The TURBO DNA-*free* ™ Kit from Ambion® (Ambion Inc, Austin, Texas, USA) was used to DNase treat all Total RNA samples following the RNeasy® RNA extraction procedure. TURBO DNase is a recombinant, engineered form of DNase I that is much more efficient than wild type DNaseI in digesting away trace amounts of unwanted DNA. TURBO DNase binds DNA substrates 6-fold more tightly than traditional DNase I, making this enzyme the tool of choice for clearing residual DNA that can generate a false positive signal in RT-PCR applications.

TURBO DNase buffer and TURBO DNase inactivation agent were thawed at room temperature prior to use. To 50µl of RNA sample, 5µl of TURBO DNase buffer was added. Following the addition of 2µl TURBO DNase, samples were incubated for 30 minutes at 37°C in a pre-warmed heat block. At 30 minutes, 5µl of TURBO DNase inactivation agent was added and incubated at room temperature for 2 minutes with occasional mixing by pipetting. Samples were centrifuged for 1.5 minutes at 10,000g. The supernatant was transferred to a new 1.5ml tube, labelled and placed on ice prior to quantity and quality assessments. Alternatively, the DNase treated RNA was stored at -80°C for future use.

### 2.3.3 Quantitation of RNA samples

The concentration of RNA was determined by measuring the absorbance at 260nm (A260) in a UV spectrophotometer Beckman DU Series 500. An absorbance of 1 unit at 260nm corresponds to 40µg of single stranded RNA per ml. 100µl of RNase-free water was added to the cuvette to blank the spectrophotometer. A 1/50 dilution of total RNA sample was made up to 100µl in RNase-free water and readings were obtained at 260nm and 280nm.

The RNA concentration was calculated as follows:

Volume of RNA sample = 50µl
Dilution = 2µl of RNA sample + 98µl RNase-free water (1 in 50 dilution)
Concentration of RNA sample = 40 x A260 x 50 (dilution factor)

### 2.3.4 Determination of Quality of total RNA

#### *2.3.4.1 Spectrophotometer, Beckman DU Series 500:*

The ratio of the readings at 260nm and 280nm provides an estimate of the purity of RNA with respect to contaminants that absorb in the UV, such as protein. Pure RNA has an A260/A280 ratio of 1.9 – 2.1 in 10mM Tris Cl, pH 7.5.

#### *2.3.4.2 Gel electrophoresis:*

RNA samples may be run on a 1% agarose gel. To make the gel, 0.5g Agarose I powder was added to 50ml 1X TBE (Tris Borate-EDTA) (Sigma-Aldrich, Dublin, Ireland) and dissolved by heating in a microwave. 3µl of Ethidium bromide (10mg/ml) was added. The solution was allowed to cool slightly before adding to appropriate gel tray and allowed to set with comb in place. When the gel was set, it was placed in the gel tank containing 1X TBE electrophoresis buffer where the samples were added. 500ng of RNA was made up to 9µl with RNase-free water. 1µl of 6X gel loading dye (Sigma-Aldrich, Dublin, Ireland) was added. Samples were run at approximately 89mV for 30 minutes. Distinct bands corresponding to 28S and 18S ribosomal RNA bands were indicative of good quality RNA.

#### *2.3.4.3 Agilent 2100 Bioanalyser:*

For microarray analysis the quality and concentration of RNA is crucial to the success of a microarray experiment. Traditional methods of assessing these include spectrophotometry and gel electrophoresis as described above, however it is now recommended that a superior method of assessing RNA quality and quantity using an Agilent 2100 Bioanalyser (Agilent Technologies, Santa Clara, CA, USA) with the RNA 6000 Nano Assay Kit (Agilent Technologies, Santa Clara, CA, USA) is employed for all gene expression microarray studies. The Agilent 2100 bioanalyser (Figure 2.9) is a highly successful microfluidics-based platform for the analysis of DNA, RNA, proteins and cells. As the first commercial analytical instrument based on lab-on-a-chip technology, the Agilent 2100 bioanalyser delivers fast,

automated, high quality digital data. All RNA samples were run on the Agilent 2100 Bioanalyser to confirm previous findings from gel electrophoresis and spectrophotometry. 1µl of sample was required. 12 samples can be run per chip with an analysis run-time of 30 minutes.

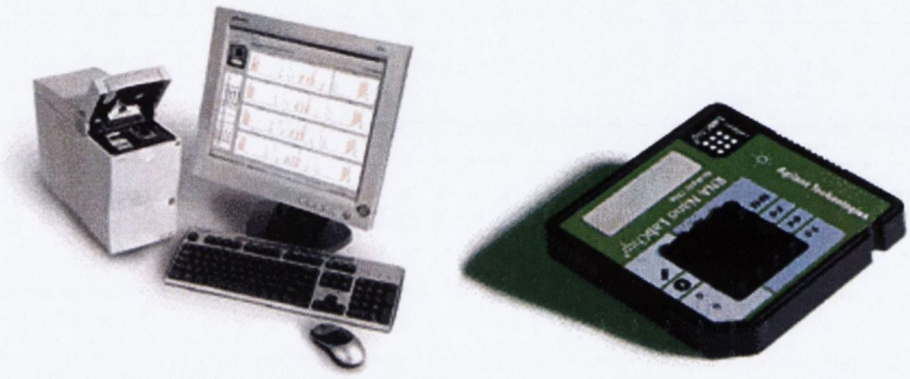


**Figure 2.9: Agilent 2100 bioanalyser and RNA 6000 Nano LabChip (Taken from Agilent 2100 Bioanalyser Expert User's Guide, May 2005).**

The RNA 6000 Nano LabChip kit provides information on the quality and quantity of RNA samples. The assay kit is designed primarily to check the quality and determine the concentration of total RNA and mRNA samples. For the analysis and quantitation of RNA samples, the RNA assay kit is useful for the analysis of samples prior to their use in more sensitive techniques and experiments such as:

- Total and mRNA purification
- Northern blot analysis
- Analysis of RT-PCR products
- Expression profiling using DNA arrays
- In-vitro RNA transcription assays
- Preparation of cDNA libraries

The RNA 6000 Nano LabChip kit has become the industry standard for the quality control and quantitation of total and messenger RNA samples. It contains a lower marker, which allows for sample alignment and permits easy comparison of samples to enable the user to distinguish

different types of mRNA based on the electrophoretic traces. Most importantly, the RNA 6000 Nano LabChip kit provides enhanced RNA quantitation simultaneous with the assessment of RNA quality thus getting both quality control as well as a good concentration measurement in one single analysis. Relative standard deviations of 10% can be achieved, while the method is much less susceptible than UV to sample contaminants such as Phenol or genomic DNA. The RNA 6000 Nano LabChip kit allows detection of total RNA down to 5ng with greater speed and accuracy than other methods. Digital data management and a series of software features are further benefits to using the Agilent 2100 bioanalyser for the analysis of RNA samples.

Due to the omnipresence of RNases and the instability of RNA, integrity checks and sample quantitation are essential steps before any assay. Microfluidics technology makes it easier and faster to analyze samples. Using as little as 5ng of total RNA (RNA nano assay) or 50 pg of total RNA (using RNA Pico for microdissected samples), the system completes the analysis unattended. The 2100 bioanalyser expert software generates the RNA Integrity Number (RIN), a quantitation estimate, and will calculate ribosomal ratios of the total RNA sample (Schroeder et al., 2006). An electropherogram and virtual gel electrophoresis is also produced for each sample. For mRNA samples, the percentage of ribosomal impurities is calculated. RIN can be used as a standard integrity measure that is independent of concentration, instrument or analyst. A RIN of greater than 5 is good total RNA quality and greater than 8 is perfect total RNA quality for downstream application (Fleige et al., 2006).

RNA quality was assessed using the RNA 6000 Nano LabChip® Kit in conjunction with Agilent 2100 Bioanalyser (Agilent Technologies, Santa Clara, CA, USA).

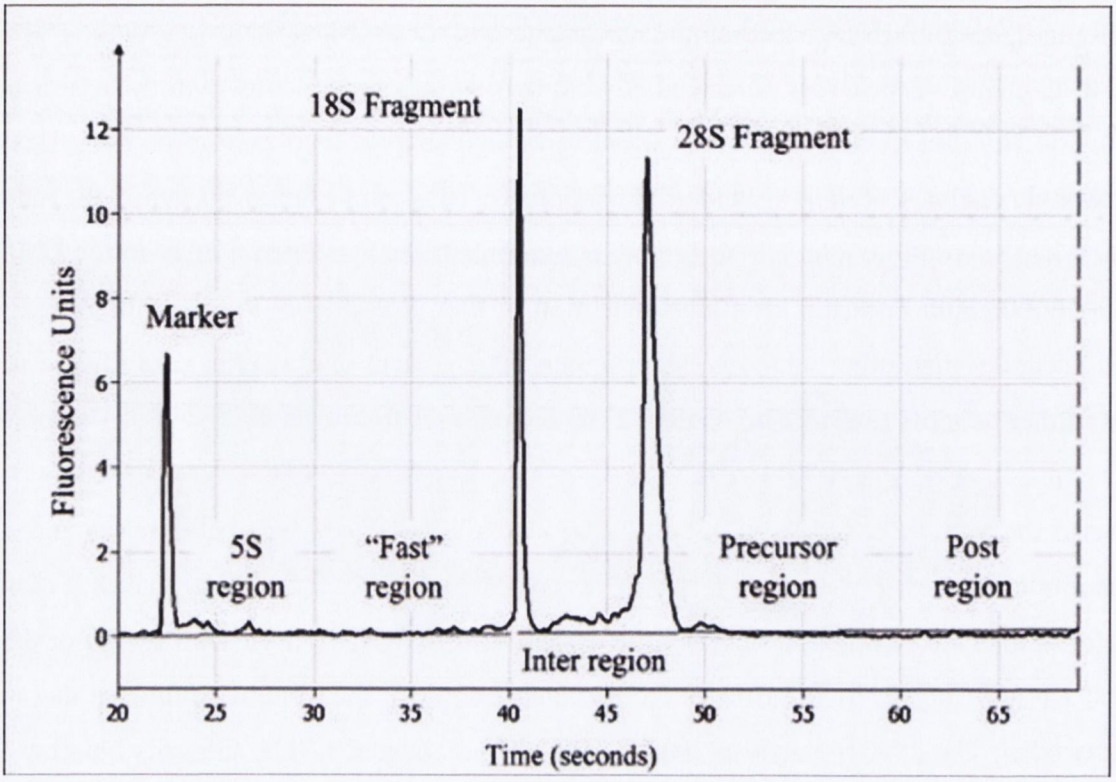



**Figure 2.10: Example of electropherogram and expected pattern from a good quality RNA sample (Taken from Agilent 2100 Bioanalyser Expert User's Guide, May 2005).**

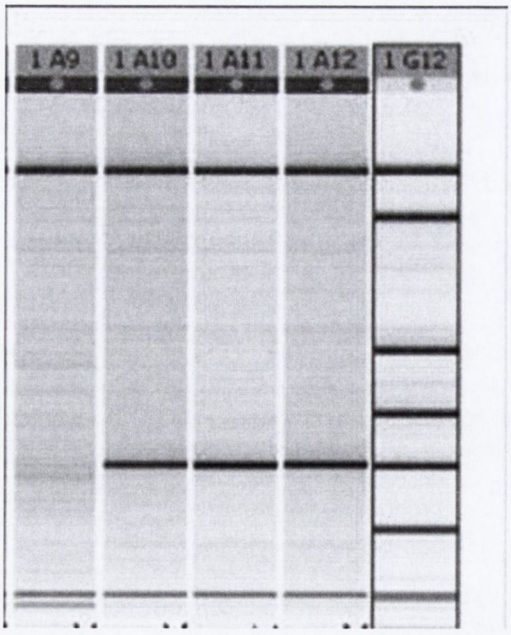



**Figure 2.11: Image of gel electrophoresis produced by the Agilent 2100 Bioanalyser (Taken from Agilent 2100 Bioanalyser Expert User's Guide, May 2005).**

## 2.4: Reverse transcription PCR

### 2.4.1 RT reaction

2μg of total RNA was used to produce cDNA for real time PCR analysis. Briefly, total RNA (2μg) in 7μl of nuclease free water and 1μl of oligo (dt) primer were added to a 0.2ml PCR microtube. This mixture was incubated at 65°C for ten minutes and cooled on ice subsequently. The mixture was spun down quickly to get any evaporation back in solution. 12μl of RT mastermix (2μl 10X RT buffer, 8μl 10mM dNTPs, 1μl 20U RNasin and 1μl (50U) MuLVRT per reaction – all obtained from Applied Biosystems, Foster City, CA, USA) was added to each 8μl RNA sample giving a total sample volume of 20μl. When making up mastermix, 10% extra volume was allowed for any loss of reagents during transfer. This mixture was incubated at 42°C for one hour and stored at 4°C short term or -20°C long term.

### 2.4.2 RT-PCR

**Table 2.7: Components and volumes per reaction for RT-PCR analysis (All components obtained from Applied Biosystems, Foster City, CA, USA). (IU = international units)**

| Components | Volume per reaction (μl) |
|---|---|
| Water | 14.875 |
| 10x buffer | 2.5 |
| 5mM $MgCl_2$ | 2.5 |
| dNTPs (10mM) | 2 |
| Primer 1 (2.5mM) | 0.5 |
| Primer 2 (2.5mM) | 0.5 |
| Taq polymerase (5IU) | 0.125 |
| ***Total*** | ***23*** |

2µl of template cDNA was added to 23µl of mastermix to give a total reaction volume of 25µl. With each sample, a positive control (2µl genomic DNA supplied by Applied Biosystems) and a negative control (2µl nuclease-free water) were also run.

The cycling conditions were as follows:

- 95°C x 5 minutes
- 25 cycles of:
    - 95°C x 30s
    - 60°C x 15s
    - 72°C x 60s
- 72°C x 7 minutes.
- Hold at 4°C.

Primers arrived as forward and reverse primers as 10nmole powders in a vial. A stock was made of each primer of 100µM (100µl of nuclease-free water added to the primer to give a 100µM stock). This was stored at -20°C. To minimise freeze thawing of the stock, a working stock of 20µM was made for each primer set - 20µl forward primer, 20µl reverse primer and 60µl or nuclease-free water giving a total working stock volume of 100µl and 20µM concentration. This was also stored at -20°C.

**Table 2.8: Primers used in RT-PCR analysis. β actin was used as a control. All primers were purchased from Applied Biosystems.**

| Markers of undifferentiated state | | Markers of differentiation | |
|---|---|---|---|
| 1. | Oct4 | 1. | Afp |
| 2. | Sox2 | 2. | Ncam |
| 3. | Nanog | 3. | Vegfr2 |

### 2.4.3 Gel electrophoresis

PCR products were visualised on 2% agarose gels (100ml 1XTBE, 2g Agarose 1 stained with 3µl of Ethidium bromide (EtBr) (10mg/ml)). Gels were made with samples run, separated and visualised on a UV transluminator.

## 2.5: Expression Array Analysis

### 2.5.1 Overview of microarray workflow and description of Applied Biosystems microarray platform.

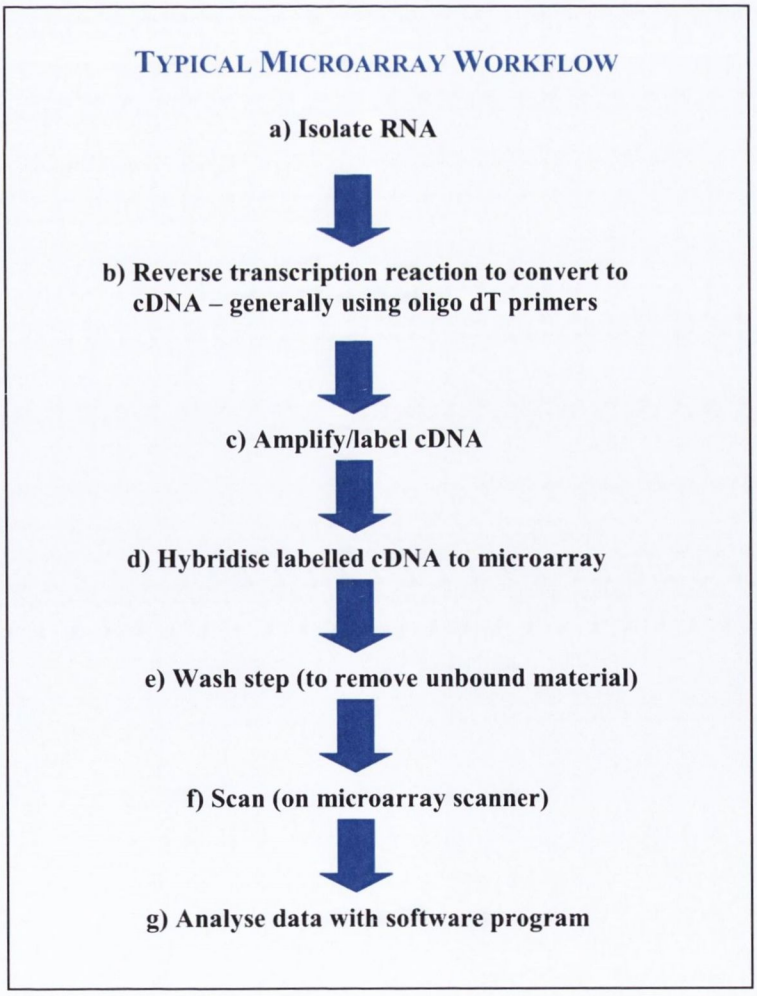



**Figure 2.12: Outline of the many steps involved in a typical microarray experiment.**

The Applied Biosystems Mouse Genome Survey Microarray, a component of the Applied Biosystems Gene Expression Array System, contains probes representing a complete, annotated, and fully curated set of approximately 32,000 mouse genes from the public and Celera databases. The Applied Biosystems Human Genome Survey Microarray provides 31,700 probes for interrogation of 27,868 genes.

- A team of experts fully curates the gene sequence of every probe.
- To ensure the highest quality and specificity, each probe is tested by mass spectrometry.
- It includes complete, searchable Celera annotations for the transcript sequences interrogated by every probe.

High-signal strength and safe, robust protocols make chemiluminescence ideal for visualizing and quantitating microarray hybridization. By combining chemiluminescence with low-background substrates, the Mouse Genome Survey Microarray ensures high signal-to-noise values and low detection limits for the entire Expression Array System. To monitor the performance of each experiment, each microarray comes with a complete set of controls. Feature controls allow normalisation of the signal variability between arrays. Chemiluminescent controls allow the monitoring of the chemiluminescent reaction. Hybridization-kit controls allow the power to monitor mixing, stringency, and washing during array hybridization. With the Expression Array System, all system components - microarray, analyzer, software, reagents, and Oracle® database - are fully integrated and optimised. This system also integrates with Applied Biosystems Real-Time PCR-based products.

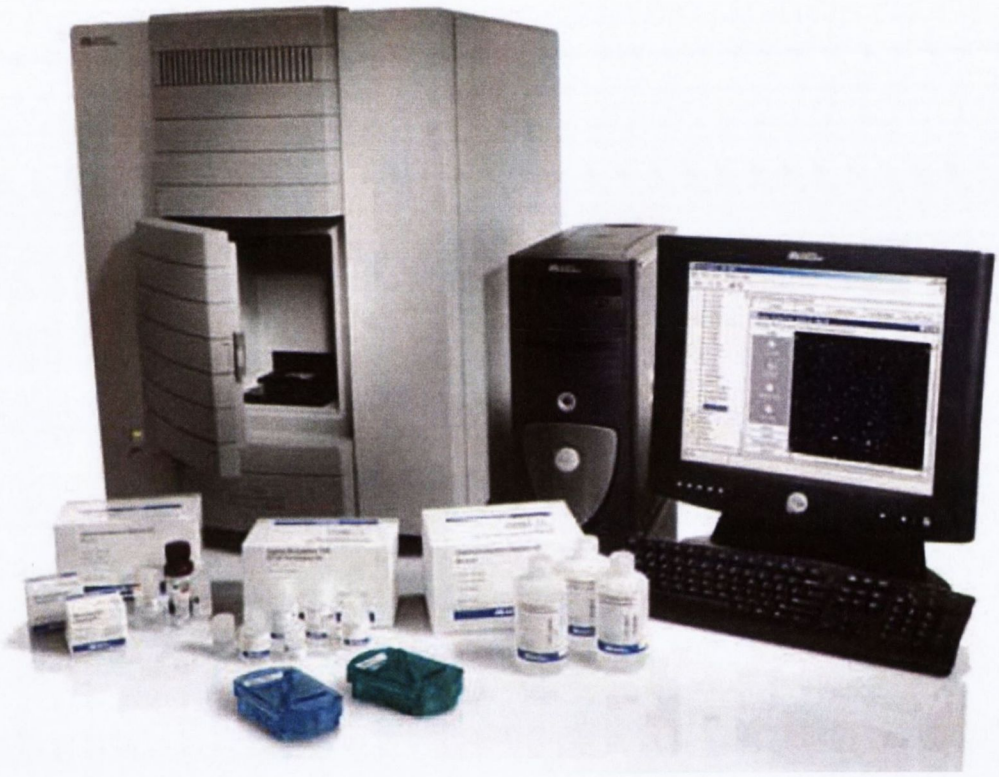


**Figure 2.13: Applied Biosystems 1700 Chemiluminescent Microarray Analyser (Taken from Applied Biosystems 1700 Chemiluminescent Microarray Analyser Chemistry Guide, 2004).**

### 2.5.2 Chemiluminescent RT-IVT labelling

The protocol for the Applied Biosystems Chemiluminescent RT-IVT labelling kit V2.0 was used to perform reverse transcription and in vitro labelling of total RNA samples.

**Step 1:** Reverse transcription: 2µg of total RNA was made up to 10µl with nuclease-free water in a 0.2ml MicroAmp tube. Added to this was, 2µl of T7-Oligo (dT) primer and 4µl of control RNA bringing the sample volume to 16µl. The samples were placed in a thermal cycler and incubated for 5 minutes at 70°C and held at 4°C. After the run, samples were placed on ice. The following components were then added to the sample in the same order: 1µl RNase inhibitor V2.0, 2µl 10x 1$^{st}$ strand buffer mix and 1µl RT enzyme V2.0 and mixed thoroughly by pipetting. Reverse transcription was then performed in a 9600 thermal cycler under the following conditions: 25°C for 10 minutes, 42°C for 2 hours, 70°C for 15 minutes and held at 4°C. After the run, tubes were placed on ice.

**Step 2:** Second strand synthesis: On ice, 55µl of nuclease-free water, 20µl 5X 2$^{nd}$ strand buffer mix V2.0 and 5µl 2$^{nd}$ strand enzyme mix were added to the 20µl sample of cDNA. Second strand synthesis was performed in a 9600 thermal cycler at 16°C for 2 hours, 70°C for 15 minutes and held at 4°C. Tubes were placed on ice after the run.

**Step 3:** cDNA Purification: In a new 1.5ml tube not supplied in the kit, 100µl of DNA binding buffer and 100µl or the entire 2$^{nd}$ strand synthesis reaction were combined. This was then added to a nano purification column placed in 2ml receptacle tube. This was centrifuged at 13000g for 1 minute. The flow through was discarded and 300µl of DNA wash buffer was added to the column. The column was centrifuged at 13000g for 1 minute. Again, the flow through was discarded and a further 300µl of DNA wash buffer was added. The column was centrifuged at 13000g for 1 minute. Flow through was discarded and the column was centrifuged again at 13000g for 1 minute. The column was then transferred to a new 1.5ml elution tube and 10µl of DNA elution buffer was added onto the fibre matrix at the bottom of the column. The column was incubated at room temperature for 1 minute prior to centrifuging at 13000g for 1 minute. A further 10µl of DNA elution buffer was added and incubated at

room temperature for 1 minute. The column was centrifuged at 13000g for 1 minute for a final elution volume of 20µl.

**Step 4:** In Vitro Transcription (IVT): The IVT components (8µl of 5X IVT buffer mix, 4µl 3.5mM DIG-UTP (Catalogue number 03 359 247 910, Roche Diagnostics GmbH, Mannheim, Germany) and 4µl of IVT enzyme mix) were added to the ds cDNA output from step 3 (brought up to 24µl with nuclease-free water) at room temperature giving a total volume of 40µl. This mixture was placed in the 9600 thermal cycler at 37°C for 9 hours and held at 4°C.

**Step 5:** Purifying cRNA: In a new 1.5ml nuclease-free microcentrifuge tube, the entire 40µl IVT reaction was combined with 20µl nuclease-free water. To this, 200µl of RNA binding buffer and 140µl 100% ethanol were added and mixed by pipetting. A RNA purification column was inserted into a 2.0ml receptacle tube and the IVT reaction-RNA binding buffer-ethanol mixture was added to the column and tube closed. The column was centrifuged at 13,000g for 1 minute. If the entire volume did not pass through the column, the centrifuge step was repeated. The column was removed and liquid discarded. The column was reinserted into the tube. The cRNA was now washed by adding 500µl of RNA wash buffer to the column. The column was centrifuged at 13,000g for 1 minute. The column was removed and liquid discarded with the column reinserted into the tube. This wash step was repeated a second time. Following the second wash, the column was again centrifuged at 13,000g for 1 minute. The column was transferred to a new 1.5ml elution tube and the cRNA was now eluted by pipetting 50µl of RNA elution buffer onto the fibre matrix at the bottom of the column. It was incubated at room temperature for 2 minutes. The column was then centrifuged at 13,000g for 1 minute to produce an elution volume of 50µl. This step was repeated to produce a total elution volume of 100µl. The column was then discarded. The cRNA product was stored on ice while the quantity and quality of the product was assessed using gel electrophoresis (see section 2.3.3 and section 2.3.4). The cRNA product can be stored for up to 2 months at -15 to -25°C or long term at -80°C.

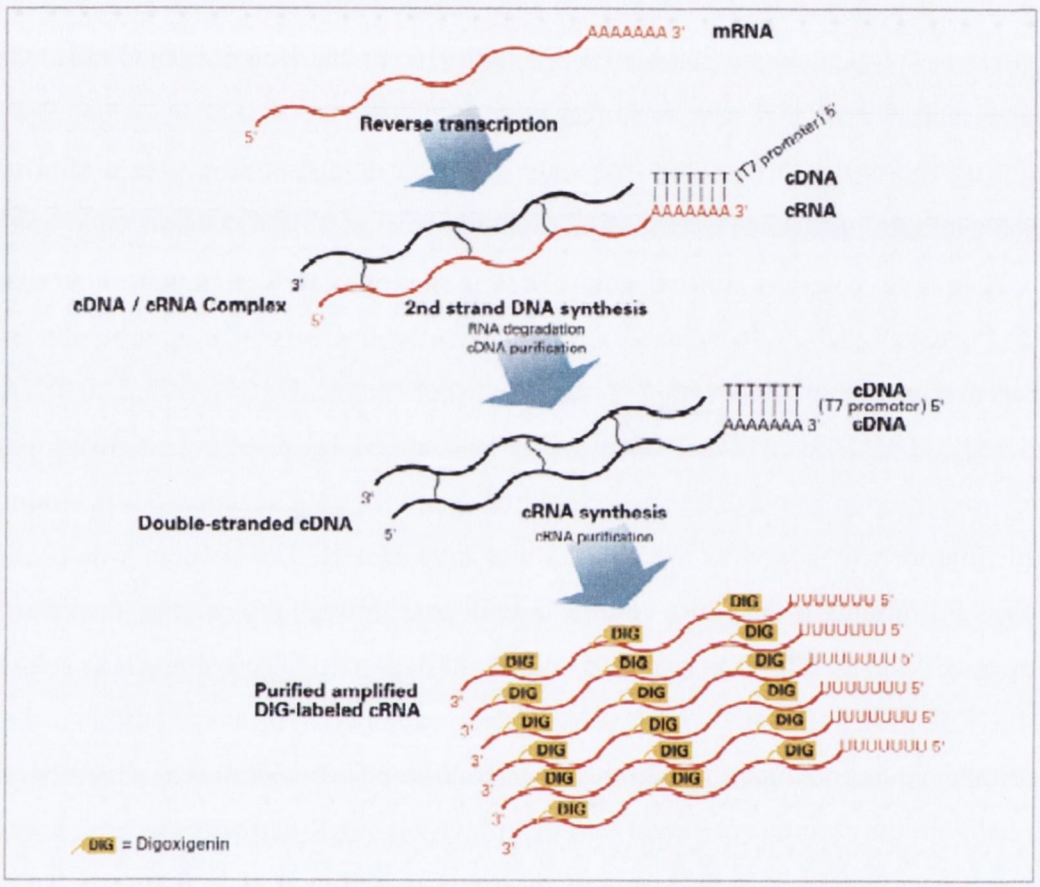



**Figure 2.14: Schematic of the RT-IVT labelling process described in Step 1 – 5 (Taken from Applied Biosystems Chemiluminescent RT-IVT Labeling Kit Protocol).**

### 2.5.3 Hybridisation and Chemiluminescent detection

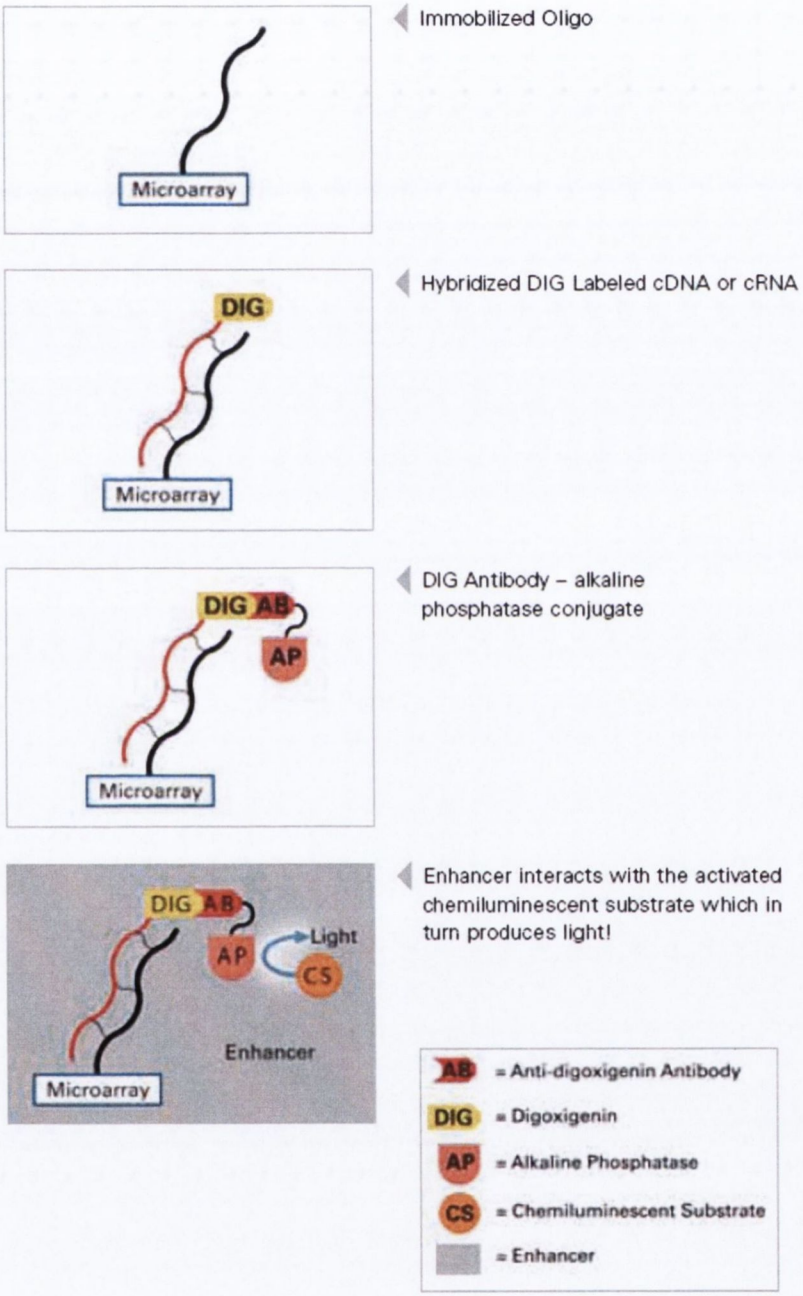



**Figure 2.15: Schematic of labelling and chemiluminescence detection step used in the expression array system microarray assay (Taken from www.appliedbiosystems.com).**

The Applied Biosystems Mouse Genome Survey Array contains approximately 34000 60-mer oligonucleotides which contain a set of approximately 1000 controls. These controls track

system performance throughout the experiment and are included in each assay. The 32,996 probes on the microarray represent 32,381 curated genes that target 44,498 transcripts. The set is compiled from publicly curated genes identified by the Mouse Genome Sequencing Consortiums, as well as genes from the Celera Genomics database. The 60-mer oligo probes are synthesised using standard phosphoramidite chemistry and solid-phase synthesis and quality controlled by mass spectrometry. The probes are deposited and covalently bound onto a derivatised nylon substrate that are backed by a glass slides by contacting spotting with a feature diameter of 180um and space of >45μm between each feature. A 24-mer oligo internal control probe (ICP) are co-spotted at every feature with 60-mer gene expression probe on the microarray.

#### *2.5.3.1 Hybridisation*

Digoxigenin-UTP labelled cRNA was generated as described in section 2.5.2 above, was hybridised to the AB Mouse genome array using the AB Chemiluminescence detection kit (Applied Biosystems, Foster City, CA 94404, USA).

Prehybridisation: Briefly, for each microarray the following pre-hybridisation mixture was prepared in a nuclease free tube, 150μl of nuclease free water, 330μl of hybridisation buffer, 100μl of hybridisation denaturant, and 420μl of blocking reagent. 1ml of the pre-hybridisation mixture was added to each microarray and incubated in a preheated hybridisation oven at 55°C for 1 hour.

Fragmenting cRNA: 10μg of labelled cRNA targets were first fragmented into 100-400 bases, as follows 10μl of cRNA fragmentation buffer, and 90μl of 10μg of DIG labelled cRNA and nuclease free water were added on ice to a 0.2ml PCR microtube. The mixture was incubated at 60°C for 30 minutes on a 9600 thermal cycler. Once complete, 50μl of cRNA stop buffer was added mixed by pipetting and the tubes were placed on ice until required.

Hybridisation of samples to array: For hybridisation of each microarray, the following components were added in a nuclease free microcentrifuge tube, 100μl of nuclease free water, 170μl of pre-warmed hybridisation buffer, 30μl of hybridisation controls ICT, 24-mer oligo

labelled with LIZ fluorescent dye), 150µl of fragmented DIG labelled cRNA, and 50µl of pre-warmed hybridisation denaturant. The mixture was vortexed to mix and centrifuged briefly.

Removing one microarray from the oven at a time, the plug was opened and 500µl of hybridisation mixture was carefully added into the port. The port was dried with lint free tissue and the cartridge resealed. The microarray cartridge was returned to the oven and incubated at 55°C at 100rpm agitation for 16 hours. It was important not to vary the hybridisation time.

After hybridisation, the following steps took place on a rocking platform shaker (tilt - angle 10°, tilt speed - 30 tilts back and forth per minute):

- Hybridisation washes:

The arrays were removed from their cartridges and the excess liquid discarded one by one. The arrays were placed in a wash tray (holds 4 arrays) and submerged in 300ml of hybridisation wash buffer 1 (30ml hybridisation wash buffer concentrate, 60ml hybridisation wash detergent concentrate, 210ml nuclease-free water). The wash trays containing 4 arrays each were placed on the rocker, vertically arranged, and agitated for 5 minutes. The buffer was then discarded. This step was repeated using 300ml of hybridisation wash buffer 2 (1.5ml hybridisation wash buffer concentrate, 298.5ml nuclease-free water). Following this two chemiluminescent rinses were performed. 300ml of chemiluminescent rinse buffer (15ml chemiluminescence rinse buffer concentrate, 285ml) was used per wash tray (4 arrays) and agitated for 5 minutes and repeated a second time.

- Antibody binding:

Following the removal of the chemiluminescence rinse buffer, 4ml of chemiluminescent blocking buffer/antibody mixture (2.8ml nuclease-free water, 0.2ml chemiluminescence rinse buffer concentrate, 1ml blocking reagent (pre heated to 37°C for 30 minutes and allowed to cool to room temperature), 15µl anti-digoxigenin-AP (Catalogue number 11093274001, Roche Diagnostics GmbH, Mannheim, Germany)) was added to each array. The arrays were then covered and put on the rocking platform for 20 minutes at room temperature.

- Antibody washes:

Three washes of 10 minutes duration each were performed using the 300ml of chemiluminescence rinse buffer (components as before) per wash tray following the removal of all liquid from the arrays.

- Chemiluminescent reaction:

A chemiluminescence enhancing rinse was performed next with 300ml of chemiluminescence enhancing rinse buffer (7.5ml chemiluminescence enhancing rinse concentrate, 292.5ml nuclease-free water) added to the tray and the wash tray placed on the rocking platform for 10 minutes. The next step was to add 4ml of chemiluminescence enhancing solution (pre heated to 37°C for 30 minutes and allowed to cool to room temperature) to each array and agitate on the rocking platform for 20 minutes. Following this, the liquid was discarded from the arrays which were then placed in 300ml of chemiluminescence enhancing rinse buffer (components as above) and agitated for 5 minutes. The last step prior to placing the array into the Applied Biosystems 1700 chemiluminescent analyser was to add 3.5ml of chemiluminescence substrate to the array. This should only be done immediately prior to placing the array into the instrument to be analysed.

#### *2.5.3.2 Chemiluminescent detection*

Eight images were collected for each microarray using the 1700 analyser which is equipped with high-resolution, large-format charge coupled device (CCD) camera, including 2 'short' chemiluminescent images (5 seconds exposure length each) and 2 'long' chemiluminescent images (25 seconds exposure length each) for gene expression analysis, 2 fluorescent images for feature finding and spot normalisation and 3 quality control (QC) images for spectrum cross-talk correction. Images were auto-gridded and the chemiluminescent signals were quantified, corrected for background and spot and spatially normalised (Stefano et al., 2005). Approximately 1,000 controls are on the microarray and are listed in Table 2.9.

**Table 2.9: Microarray controls.**

| Microarray Controls | | |
|---|---|---|
| **Control Type** | **No of features** | **Function** |
| Assay background controls | 196 | Quality check for assay background |
| Blank features | 80 | Prevent cross-talk |
| Control ladder (CL) | 20 | Positive control for chemiluminescent detection chemistry across the dynamic range |
| Control ladder (FL) | 20 | Positive fluorescence control |
| Hybridisation controls | 120 | Quality check for hybridisation reaction |
| IVT labelling controls | 120 | Quality check for in vitro transcription in RT-IVT labelling kit |
| Landmark fiducials (CL) | 96 | System software uses to grid array during image analysis |
| Landmark fiducials (FL) | 16 | System software uses to grid array during image analysis |
| RT labelling controls | 120 | Quality check for reverse transcription in RT and RT-IVT labelling kits |
| Spatial calibration controls | 117 | Normalisation of chemiluminescent and fluorescent images across the array |
| Manufacturing QC controls | 442 | Used in manufacturing as a quality check |

### 2.5.4 Microarray analysis

After capturing array images and quantification of chemiluminescence for each probe, the AB1700 performs initial quality control analysis. Each array must pass this quality control to be included in subsequent analysis. For the purposes of this study, microarray data was then

exported and imported into a specific ABarray package within Bioconductor R, which allows analysis of the AB1700 array data files. The Applied Biosystems Genome Survey Array Data Analysis program is also called the ab1700gui. The ab1700gui performs normalisation and quality assessment on output from the AB1700. It can also be used to perform primary analysis and output lists of significant genes. Much of this analysis is carried out by a second program called R. This program applies filtering methods to identify genes that show differential expression across two study groups. The ab1700 package for R software filters data using signal to noise (S/N) ratio threshold (default = 3). It reads the output from the AB1700 software and performs analysis on hybridisation control spike-ins, raw data QA, associated plots (boxplots for signal distribution range, MA plot for signal distribution and signal variability, CV plot for variation among hybridisation replicates, Scatter plot for correlation between hybridisation arrays, Correlation heatmap for visualisation, S/N detection concordance), quantile normalisation, repeat data QA after normalisation, performs t-test and fold change with graphics to visualise t test results. Plots will be generated for control probes but they will be excluded for further analysis. The data file and an experiment design file containing names of arrays to be analysed are uploaded to R which performs the above functions.

Names and functions of all genes of interest were downloaded from the PANTHER (**P**rotein **AN**alysis **TH**rough **E**volutionary **R**elationships) Classification System, available at www.pantherdb.org. Secondary analysis was also performed using PANTHER; the annotated gene list was analysed to identify molecular functions, biological processes and pathways that showed increased representation between the two study groups, either due to upregulation or downregulation of component genes. Cluster analysis was performed using the Cluster 3.0 and TreeView 1.04. The statistical analysis process is summarised in Figure 2.16.

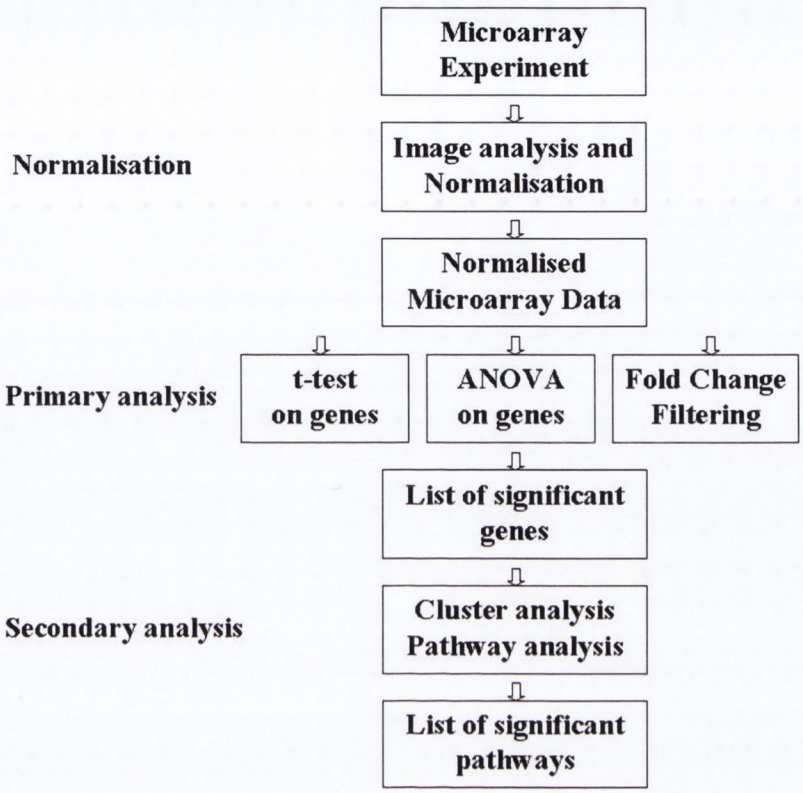



**Figure 2.16: Overview of data analysis methods.**

### *2.5.4.1 Reducing bias; normalisation and replicates*

Systematic bias is the inherent tendency of a system to favour particular outcomes or to make particular errors. Random bias is due to inaccurate results or readings that occur randomly and cannot be predicted. In the case of microarray experiments, comparing the output from two chips to which the same mRNA sample is applied reveals both systematic and random bias (Figure 2.17). All measurements from one chip should match those on the other exactly, so in Figure 2.17, all points on the graph should lie on the diagonal. Systematic bias is revealed by a deviation from the diagonal that increases with intensity – this bias is relatively predictable and is signal dependent. Normalisation is an attempt to correct for systematic bias in data.

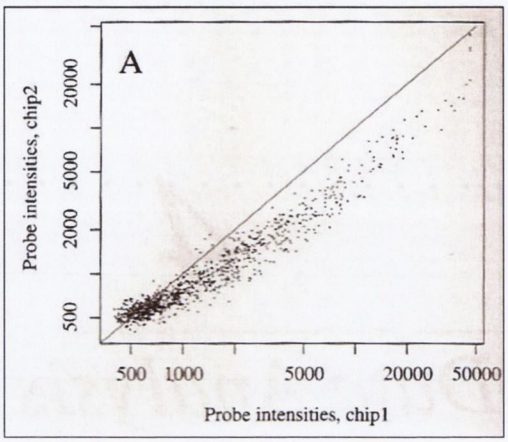



**Figure 2.17: Comparison of output from 2 microarrays to which the same mRNA sample was applied (Knudsen, 2004).**

Normalisation involves equalising data from different channels before analysis. It corrects for technical variations between and within single hybridisations, namely quantity of starting RNA and labelling and detection efficiencies for each sample. There are a variety of normalisation schemes in use; global normalisation has been applied to the data presented in subsequent chapters. Global normalisation assumes that given a large enough sample, the average signal intensity (corresponding to gene intensity level) is constant. M *versus* A plots or MA plots showing logarithm of signal ratios against logarithm of average signal intensity are often used to detect signal dependent biases and efficacy of normalisation. Effect of global normalisation on an MA plot of the data in the previous example is shown in Figure 2.18.

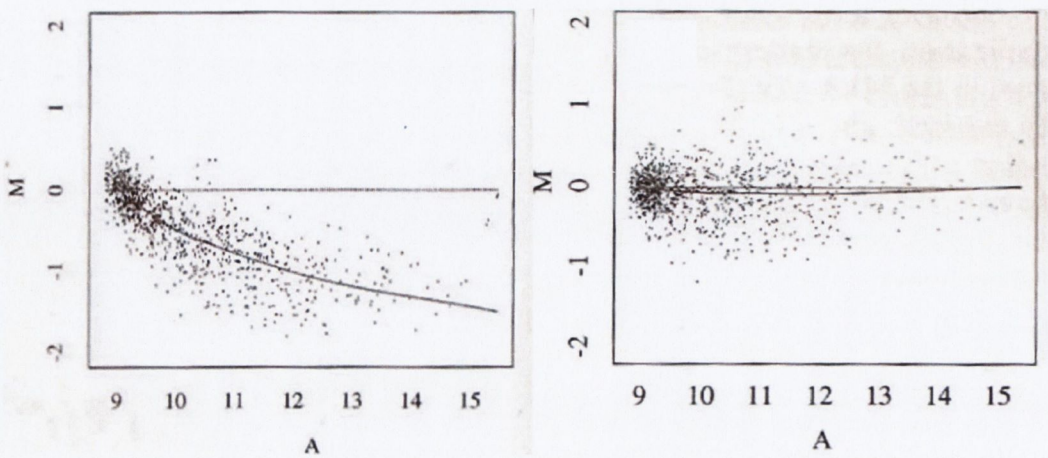



**Figure 2.18: Comparison of MA plots of data from Figure 2.17 before (left) and after (right) global normalisation. (Knudsen, 2004).**

Random bias is more unpredictable than systematic bias; this is often not a problem as random errors on average cancel each other out. One can minimise random bias by using replicates. Replicates estimate variation within the system and allow the use of t tests or ANOVA to determine whether differences between results are real, or whether they could be due to noise within the system.

Biological replicates can refer to hybridisations that involve mRNA from a different extraction or from a completely different sample. Generating replicates is both time-consuming and expensive, but the more replicates, the better the statistics. Three replicates is the recommended minimum.

Technical replicates between arrays refers to replication in which the target mRNA is from the same pool, that is, from the same extraction or RT-IVT reaction. In this study, as the biological replicates were so close, technical replicates were not used.

#### *2.5.4.2 Primary analysis using ab1700gui*

After normalisation and quality assessment, ab1700gui was used to conduct t tests and fold change analysis, looking for genes that showed significantly different levels of expression between two different cohorts (or classes). The output of this analysis was filtered to identify genes of interest.

##### *2.5.4.2.1 Class Comparison*

Class comparison is a method of data analysis that compares expression profiles from predefined specimen classes. The aims of class comparison are to determine whether the expression profiles differ among the classes and if so, to identify genes that are differentially expressed between classes. If possible, class comparison aims to develop a multivariate predictor of class membership based on level of expression of certain genes. Identification of differentially expressed genes was based on fold change analysis and significance testing.

*2.5.4.2.2 Significance*

The ab1700gui was used to perform a two-sample t test between two comparison groups. This testing was preceded by a probe filtering procedure to remove undetected probes: probes in which the signal to noise ratio was less than 3. A threshold value, usually 50%, was selected for t test analysis; if a probe was not detected in more than 50% of the samples in a class, then it was listed as being undetectable in that class. For each probe, the t test looked at the mean and the variance of expression levels in class 1 compared with those in class 2 and calculated the probability of the observed difference in means occurring when the null hypothesis is true; i.e. when the mean of the two groups was equal. The t test in ab1700gui assumes unequal variance between groups. The output of this testing was a p value for each detected probe.

*2.5.4.2.3 Correction for multiple testing*

It is important to consider the effect of multiple testing in microarray experiments. A p value of 0.05 implies a 5% probability of a false positive (type 1) error when looking at a single gene. However, the same p value implies 500 false positives when looking at 10,000 genes. To get a 5% type 1 error rate with 10,000 genes requires a p value of 0.000005. This is a very strict cut off point, which very few genes will satisfy. A compromise is using a False Discovery Rate (FDR) or adjusted p value of 5%. This implies 5 type 1 errors in a list of 100 significant genes. The ab1700gui calculates FDR values from p values using the Benjamini-Hochberg procedure (Benjamini et al., 2001). The results are written into files and plotted into MA and volcano plots.

*2.5.4.2.4 Fold Change*

Fold change (FC) and $\log_2$ (FC) for detected probes were also calculated. FC compares the expression level of the same gene in both classes and assesses how many fold up - or downregulated the gene is in one class compared with the other. In microarray studies, where number of replicates is low, but number of genes is high, there is a significant risk of experimental error. It is particularly difficult to estimate variance given a small number of

replicates. To allow for this, it is recommended that only genes with a FC ≥ 2 are included for analysis. This guards against low p values that arise from underestimation of variance (Knudsen, 2004).

*2.5.4.2.5 Gene Filtering*

Fold changes, p values from t tests and FDR values from multiple testing corrections were all used to filter gene lists from the 32,381 genes represented on each array to lists of between 100 to approximately 1000 genes that showed a significant difference in expression levels between 2 classes. Initially, basic filtering was automatically applied by ab1700gui, where flagged genes (that did not pass quality control) and those with a signal-to-noise ratio (S/N) of less than 3 were deemed undetectable and removed from further analysis. All control elements on the array such as hybridisation controls, RT-IVT controls, etc. were also removed, after being screened to ensure the data met the quality control requirements.

Preferentially, genes with an FDR ≤ 0.05 were considered to show significant differential expression. However, if no or very few genes passed the Benjamini-Hochberg corrections, those with the smallest p value were evaluated, using $p \leq 0.01$. Finally, only genes with a FC ≥ 2 were included in subsequent analysis.

***2.5.4.3 Secondary analysis using PANTHER***

PANTHER (www.pantherdb.org) is a unique resource that classifies genes by their functions, using published scientific experimental evidence as well as sequence analysis to predict function even in the absence of direct experimental evidence (Thomas et al., 2003a; Thomas et al., 2003b). Gene products are classified into families and subfamilies of shared function, which are then categorized by molecular function and biological process. For an increasing number of proteins, interactions in biochemical pathways are also characterised. A key function of the PANTHER system is the ability to analyse gene lists for over- or under-represented biological processes.

### *2.5.4.4 Cluster analysis*

Using the raw data from the AB microarray reader, cluster analysis was performed with the freely available clustering software, Cluster 3.0. Cluster 3.0 grouped together genes with comparable patterns of expression by employing mathematical similarity algorithms. The microarray data was up-loaded in a tab-delimited text (*.txt) file containing the probe IDs in the left most column followed by columns of the microarray data. To reduce background noise, the uploaded genes were filtered to remove genes with very low expression levels in all experimental conditions using the Cluster 3.0 function: *'At least X observations abs(Val) > = Y'* where 'Y' was set to >/+ '8' for each normalised probe signal. The data was adjusted by five cycles of median centralisation followed by five additional cycles of normalisation as per the Cluster 3.0 protocol using the 'Adjust data' function. Median centralisation adjusted the values in each row (probe signal) subtracting the row-wise mean from the values in each row so that median value of each row approached or was very close to zero. The normalisation step multiplied all values in each row by a calculated scale factor 'S' so that the sum of the squares of the values in each row approached or was very close to one. Following the data adjustment steps, Cluster 3.0 began the unsupervised hierarchical clustering algorithm. The initial step involved assessing all the genes in the data set without bias (unsupervised), and the two genes with the most similar expression levels were joined to create a first node. These genes were then removed from the dataset and replaced by a value that represented the new node. Subsequent nodes were determined and added by repeated cycles of pair wise joining of genes and/or nodes, based on the distance between them, culminating in all genes belonging to the one node. Very short branches joined the genes if their expression levels were similar with increasingly longer branches as their similarity decreased. The analysis generated three output files: *.cdt, *.gtr and *.atr. The *.cdt (for clustered data table) file contained the original data with the rows and columns reordered based on the clustering result. The *.gtr (gene tree) and *.atr (array tree) files are tab-delimited text files that report on the history of node joining in the gene or array clustering.

#### *2.5.4.5 TreeView 1.04*

Tree View 1.04 is a cross-platform open source Java (programming language) based software application that allows interactive graphical analysis of the results from Cluster 3.0. TreeView reads in the *.cdt, *.gtr and *.atr files generated by Cluster 3.0 analysis and plots graphical dendrograms (tree diagrams) of the clustered genes expression data. The graphic output generated by TreeView is composed of two main components: (1) the dendrogram (on the left of the image) representing the degree of correlation between genes and/or nodes and (2) adjacent to the dendrogram (on the right of the image), the gene expression data organised into rows and columns where the rows correspond to genes and the columns to experiments (microarrays). The value of each row in a column is represented in a red/green colour scale, where red indicates higher gene expression, green indicates lower gene expression and black indicates no change in gene expression in that experiment relative to the other experiments in the clustered dataset.

#### *2.5.4.6 Spotfire DecisionSite™ for Functional Genomics:*

Spotfire DecisionSite™ for Functional Genomics (Spotfire AB, Goteborg, Sweden) was used for more complex data comparisons. Differentially expressed lists of genes obtained from the ABarray output were imported into Spotfire® prior to elimination of those genes with FDR $\leq$ 0.05 and FC $\geq$ 2 for further analysis.

### 2.6: TaqMan® RT PCR validation

RNA from all samples was used for TaqMan® polymerase chain reaction (PCR) validation in a series of 47 targets (Table 2.11). RNA was reverse transcribed using a High Capacity cDNA Archive Kit (Applied Biosystems, CA, USA). Primers and probes for TaqMan® PCR were obtained by using Applied Biosystems' pre-designed TaqMan® Gene Expression Assays. PCR was carried out using an ABI PRISM 7900 Sequence Detection System (Applied Biosystems, CA, USA). Analysis of relative gene expression data was performed using the

comparative $C_T$ method with glyceraldehyde-3-phosphate dehydrogenase (GAPDH) as an endogenous control/reference assay.

### 2.6.1 Reverse transcription of total RNA to cDNA:

The High-Capacity cDNA archive kit from Applied Biosystems contains reagents for reverse transcription of total RNA (0.1µg to 10µg) to single-stranded cDNA. The kit reagents are thawed on ice prior to use.

**Step 1: Preparation of reaction master mix:**

**Table 2.10: Components and volumes per reaction for preparation of reverse transcription mastermix.**

| Component | Volume (µl)/reaction |
|---|---|
| 10X Reverse transcription buffer | 10 |
| 25X dNTPs | 4 |
| 10X random primers | 10 |
| MultiScribe™ Reverse Transcriptase, 50U/µl | 5 |
| Nuclease-free water | 21 |
| ***Total per reaction*** | ***50*** |

The number of reactions was calculated per experiment and an additional 10% volume was allowed to allow for loss during reagent transfers.

**Step 2: Preparation of RNA sample for conversion to cDNA:**

500ng of RNA from each RNA sample was used and made up to 50µl with nuclease - free water to give a total reaction volume of 100µl.

**Step 3: Performing reverse transcription:**

The 100µl reaction samples were placed in PCR tubes and capped. They were placed in the GeneAmp® 9600 Thermal cycler at the following thermal cycler conditions with the reaction volume set at 100µl:

**Table 2.11: Thermal cycler conditions for reverse transcription.**

| | Step 1 | Step 2 |
|---|---|---|
| **Temperature** | 25°C | 37°C |
| **Time** | 10 minutes | 120 minutes |

This was followed by holding at 4°C. The cDNA was then either stored for long term use in the -20°C freezer or if for immediate or up to 24 hours before use was stored at 4°C.

### 2.6.2 Quantitative Real-time RT-PCR

Quantitative real-time RT-PCR (qPCR) was commercialised in 1995 and is generally considered the gold standard assay for measuring gene expression and is often used to validate gene expression array data (Qin et al., 2006). This method capitalises on the fact on that there is a quantitative relationship between the amount of starting target sample and the amount of PCR product at any given PCR cycle number. The exonuclease activity of AmpliTaq® DNA polymerase is used with a cleavable fluorescent probe in combination with forward and reverse PCR primers. TaqMan® probes contain a reporter dye (6-FAM™ dye) linked to the 5' end of the probe, and a non-fluorescent quencher (NFQ) at the 3' end of the probe. When the probe is intact, the proximity of the reporter dye to the quencher results in suppression of the reporter fluorescence. During PCR, the TaqMan® probe anneals specifically to a complementary sequence between the forward and reverse primer sites. Only probes that are hybridised to the complementary target are cleaved by the 5' exonuclease activity of AmpliTaq® DNA polymerase. Cleavage separates the reporter dye from the NFQ, resulting in an increase in reporter dye fluorescence at each PCR cycle. This increase in fluorescent signal occurs only if the target sequence is complementary to the probe and is amplified during PCR. The amount of fluorescence produced from the TaqMan® probe is measured at each

amplification cycle, providing a look at the "real-time" changes in the amplification product as the PCR process unfolds. Identification of the PCR cycle when the exponential growth phase is first detectable (Cycle threshold, $C_T$) provides extremely accurate quantitation of gene expression in the starting samples. In fact, real-time RT-PCR with TaqMan® probes is acknowledged to be the most reliable and sensitive method of quantifying gene expression.

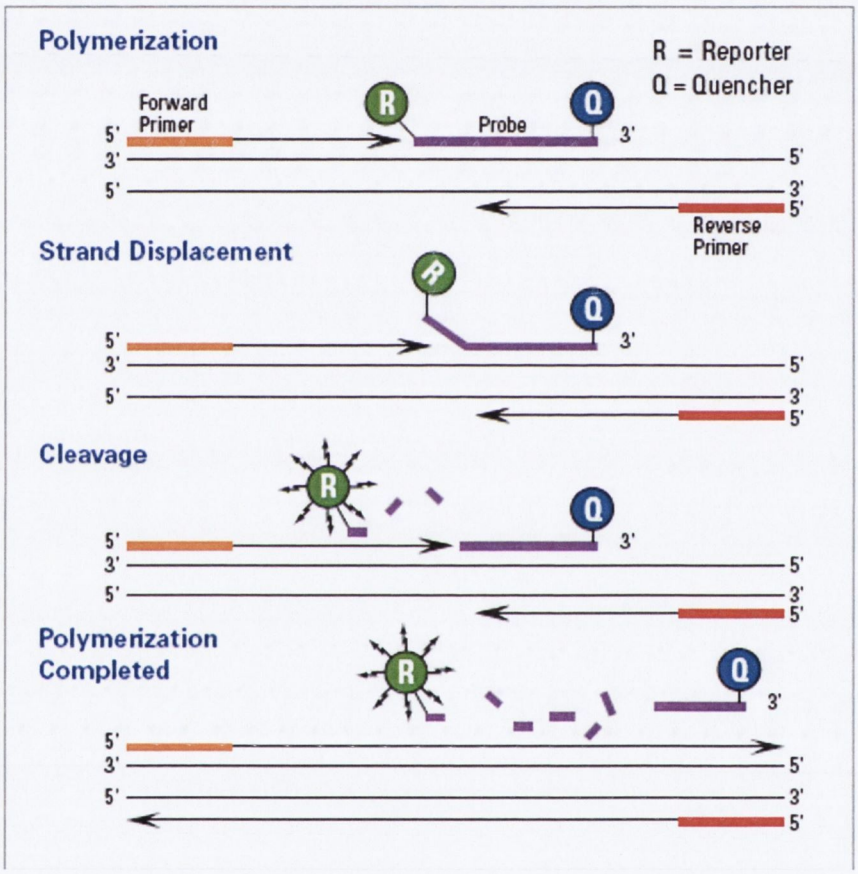



**Figure 2.19: This figure details the 5' nuclease reaction showing the mechanism of probe cleavage resulting in a fluorescent signal (Taken from TaqMan® Gene Expression Assays Protocol from Applied Biosystems).**

This method has been shown to be the most efficient and accurate method to analyse and validate microarray data (Rajeevan et al., 2001b). Methods used previously to verify expression array data included Northern blot hybridisation and RNase protection assay which require as much as 5μg of RNA. Quantitative real-time RT-PCR, however, requires less RNA (≤1μg) and is less time consuming than older methods.

Comparison of samples requires normalisation to compensate for differences in the amount of biological material present in the tested samples. Some methods used are normalising total RNA amount to ribosomal RNA, to externally added RNA standard or to internal reference genes. The latter method is the most common and is the method used here. Mouse GAPDH was used as an endogenous control and obtained from Applied Biosystems. GAPDH was used to normalise the expression levels of target genes by correcting differences in the amount of cDNA that is loaded into PCR reaction wells. GAPDH was monitored to ensure that it was consistently expressed across the sample set as it must be uniform across all samples to be an appropriate endogenous control. Finding an appropriate reference gene is widely recognised as being problematic as there is no universal reference gene with a constant expression in all tissues (Kubista et al., 2006).

**Table 2.12: Components and volume per reaction for Quantitative real-time RT-PCR. * This was not a multiplex assay but rather a separate reaction was performed using Custom designed TaqMan® Gene Expression Assays (47 assays/targets used, see table 2.13).**

| Component | Volume per reaction | |
|---|---|---|
| | 1 reaction (µL) | 1 reaction (µL) |
| Taqman® Universal PCR Master Mix, No AmpErase® UNG (2X) | 10 | 10 |
| Gapdh Reverse Primer | 0.6 | - |
| Gapdh Forward Primer | 0.6 | - |
| Gapdh probe | 0.4 | - |
| cDNA | 5 | 5 |
| Water | 3.4 | 4 |
| Assay * | - | 1 |
| **Total** | **20** | **20** |

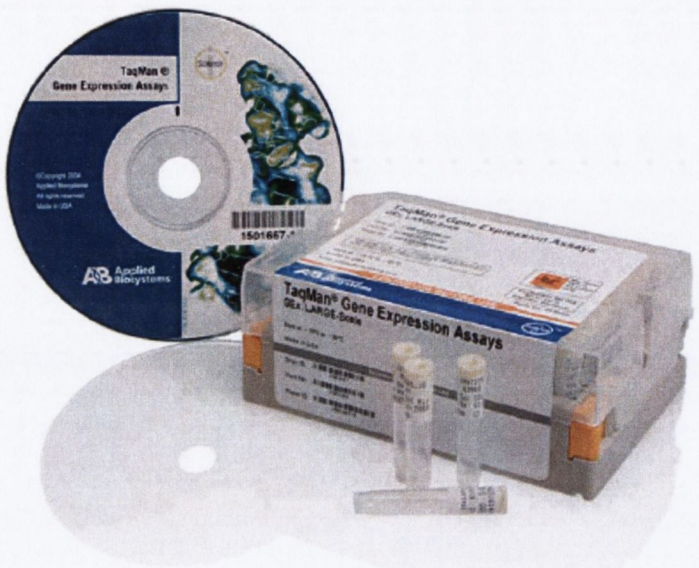



**Figure 2.20: TaqMan® Gene Expression Assays from Applied Biosystems (Taken from https://products.appliedbiosystems.com/ab/en/US/htdocs/productMgr/images/AB-1000_small.jpg).**

TaqMan® Gene Expression Assays consist of a 20X mix of unlabelled PCR primers and TaqMan® MGB probe (FAM™ dye-labelled) (Table 2.11). These assays are designed for the detection and quantitation of specific genetic sequences in RNA samples converted to cDNA.

As before, when preparing multiple reactions, a 10% excess volume was allowed for any reagent loss during reagent transfer.

96 well plates were used and all samples were run in quadruplicate. The endogenous control, mouse Gapdh (Mm99999915_g1), was run in each plate. NTCs (no template controls) in quadruplicate were run for each plate. The following thermal cycler conditions were used using the Applied Biosystems 7500 Fast Real-Time PCR system:

- Hold for 2 minutes at 50°C.
- Hold for 10 minutes at 95°C.
- 40 cycles of:

- 15 seconds at 95°C.
- 1 minute at 60°C.

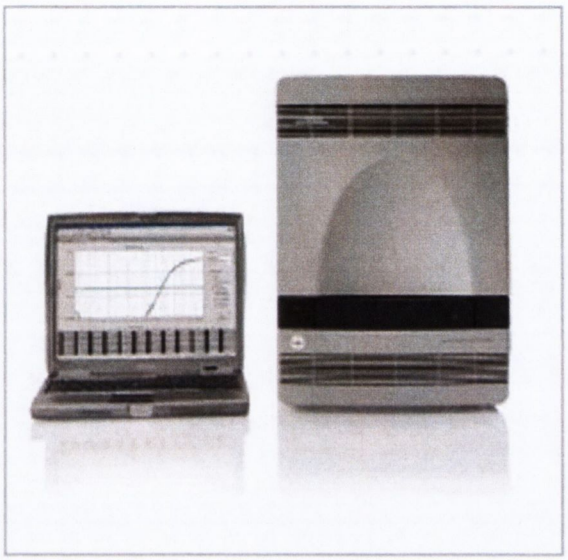


**Figure 2.21: Applied Biosystems 7500 Fast Real-Time PCR System (Taken from https://products.appliedbiosystems.com/ab/en/US/htdocs/productMgr/images/7500_Real_Time_thumb.jpg).**

The Applied Biosystems 7500 Fast Real-Time PCR system is an integrated platform for the detection and quantification of nucleic acid sequences.

## Table 2.13: List of Validated Targets Used for Quantitative Real-time PCR in mouse cell lines.

| Gene Symbol | Gene Name | Probe ID | RefSeq NM | AB Assay ID |
|---|---|---|---|---|
| *Adamts5* | A disintegrin-like and metalloprotease (reprolysin type) with thrombospondin type 1 motif, 5 (aggrecanase-2) | 429467 | NM_011782.1 | Mm00478620_m1 |
| *Afp* | alpha fetoprotein | 929310 | NM_007423.2 | Mm00431715_m1 |
| *Ccnd1* | Cyclin D1 | 684806 | NM_007631.1 | Mm00432359_m1 |
| *Cdkn1c* | cyclin-dependent kinase inhibitor 1C (P57) | 930704 | NM_009876.2 | Mm00438170_m1 |
| *Cited1* | Cbp/p300-interacting transactivator with Glu/Asp-rich carboxy-terminal domain 1 | 902496 | NM_007709.3 | Mm00455934_m1 |
| *Dcn* | Decorin | 409948 | NM_007833.1 | Mm00514535_m1 |
| *Dscam* | Down syndrome cell adhesion molecule | 721264 | NM_031174.2 | Mm00518961_m1 |
| *Egr4* | early growth response 4 | 493170 | NM_020596.1 | Mm00842279_g1 |
| *Eif2s3y* | eukaryotic translation initiation factor 2, subunit 3, structural gene Y-linked | 931120 | NM_012011.1 | Mm00468995_g1 |
| *Fgf5* | fibroblast growth factor 5 | 783155 | NM_010203.2 | Mm00438919_m1 |
| *Foxj1* | forkhead box J1 | 660384 | NM_008240.2 | Mm00807215_m1 |
| *Fzd1* | frizzled homolog 1 (Drosophila) | 435251 | NM_021457.2 | Mm00445405_s1 |
| *Galgt2* | UDP-N-acetyl-alpha-D-galactosamine: (N-acetylneuraminyl)-galactosyl –N-acetylglucosaminylpolypeptide-beta-1, 4-N-acetylgalactosaminyltransferase | 698706 | NM_008081.1 | Mm00484661_m1 |
| *Gata1* | GATA binding protein 1 | 689522 | NM_008089.1 | Mm00484678_m1 |
| *Gata6* | GATA binding protein 6 | 500605 | NM_010258.2 | Mm00802636_m1 |
| *Gdf15* | growth differentiation factor 15 | 786761 | NM_011819.1 | Mm00442228_m1 |
| *Hhip* | Hedgehog-interacting protein | 605104 | | Mm00469580_m1 |
| *Hoxb1* | homeo box B1 | 427869 | | Mm00515118_g1 |
| *Hoxd1* | homeo box D1 | 364937 | NM_010467.1 | Mm00439370_g1 |
| *Hoxd9* | homeo box D9 | 738419 | NM_013555.2 | Mm00442840_m1 |
| *Igfbp5* | insulin-like growth factor binding protein 5 | 729395 | NM_010518.1 | Mm00516037_m1 |
| *Klf2* | Kruppel-like factor 2 (lung) | 709254 | NM_008452.1 | Mm00500486_g1 |
| *Klk6* | Kallikrein 6 | 884776 | NM_010639.5 | Mm00834006_g1 |
| *Lefty1* | left right determination factor 1 | 378445 | | Mm00438615_m1 |
| *Lifr* | leukemia inhibitory factor receptor | 844166 | | Mm00442940_m1 |
| *Mafb* | v-maf musculoaponeurotic fibrosarcoma oncogene family, protein B (avian) | 373273 | NM_010658.2 | Mm00627481_s1 |
| *MMP2* | matrix metalloproteinase 2 | 871400 | NM_008610.2 | Mm00439508_m1 |
| *Myc* | myelocytomatosis oncogene | 819312 | NM_010849.2 | Mm00487803_m1 |
| *Nanog* | Nanog homeobox | 593412 | XM_132755 | Mm02019550_s1 |
| *Notch4* | Notch gene homolog 4 (Drosophila) | 341143 | NM_010244.2 | Mm00440525_m1 |
| *Olfr1450* | olfactory receptor 1450 | 652688 | NM_146371.1 | Mm00526458_s1 |
| *Olig3* | oligodendrocyte transcription factor 3 | 420862 | NM_053008.1 | Mm02525085_s1 |
| *Pou5f1* | POU domain, class 5, transcription factor 1 | 615439 | NM_013633.1 | Mm00658129_gh |
| *Pramel6* | preferentially expressed antigen in melanoma like 6 | 590714 | NM_178249.2 | Mm00626457_m1 |
| *Rab15* | RAB15, member RAS oncogene family | 917261 | NM_134050.2 | Mm00513627_m1 |
| *Rhoj* | ras homolog gene family, member J | 678861 | NM_023275.1 | Mm00502666_m1 |
| *Scml2* | sex comb on midleg-like 2 (Drosophila) | 651398 | NM_133194.2 | Mm00453761_m1 |
| *Slamf9* | SLAM family member 9 | 774433 | NM_029612.2 | Mm00504048_m1 |
| *Sox17* | SRY-box containing gene 17 | 774026 | NM_011441.2 | Mm00488363_m1 |
| *Sox2* | SRY-box containing gene 2 | 848340 | NM_011443.2 | Mm00488369_s1 |
| *Tcstv1* | 2-cell-stage, variable group, member 1 | 801958 | NM_018756.2 | Mm02525932_s1 |
| *Tert* | telomerase reverse transcriptase | 568446 | NM_009354.1 | Mm00436931_m1 |
| *Tex12* | testis expressed gene 12 | 919185 | NM_025687.1 | Mm00499441_m1 |
| *Timp2* | tissue inhibitor of metalloproteinase 2 | 908795 | NM_011594.2 | Mm00441825_m1 |

| | | | | |
|---|---|---|---|---|
| *Tnc* | tenascin C | 915023 | NM_011607.1 | Mm00495662_m1 |
| *Tnfrsf9* | tumor necrosis factor receptor superfamily, member 9 | 461834 | NM_011612.1 | Mm00440525_m1 |
| *Wnt6* | wingless-related MMTV integration site 6 | 590115 | NM_009526.2 | Mm00437351_m1 |

**Table 2.14: List of Validated Targets Used for Quantitative Real-time PCR in human cell lines.**

| Gene Symbol | Gene Name | Probe ID | RefSeq NM | AB Assay ID |
|---|---|---|---|---|
| *Akt3* | v-akt murine thymoma viral oncogene homolog 3 protein kinase B, gamma) | 190556 | NM_005465.3 | Hs00178533_m1 |
| *Chn2* | Chimerin (chimaerin) 2 | 113377 | NM_004067.1 | Hs00187896_m1 |
| *Dbc1* | Deleted in bladder cancer 1 | 125479 | NM_014618.1 | Hs00180893_m1 |
| *Dlx5* | Distal-less homeo box 5 | 173146 | NM_005221.4 | Hs00193291_m1 |
| *Eno3* | Enolase 3 (beta, muscle) | 211773 | NM_053013.1 | Hs00266551_m1 |
| *Fos* | v-fos FBJ murine osteosarcoma viral oncogene homolog | 205128 | NM_005252.2 | Hs00170630_m1 |
| *Fzd4* | frizzled homolog 4 (Drosophila) | 169119 | NM_012193.2 | Hs00201853_m1 |
| *Fzd6* | frizzled homolog 6 (Drosophila) | 118650 | NM_003506.2 | Hs00171574_m1 |
| *Gpsm1* | G-protein signalling modulator 1 (AGS3-like, C.elegans) | 142837 | | Hs00293424_m1 |
| *HesX1* | Homeo box (expressed in ES cells) 1 | 207047 | NM_003865.1 | Hs00172696_m1 |
| *Jun* | v-jun sarcoma virus 17 oncogene homolog (avian) | 123273 | NM_002228.3 | Hs00277190_s1 |
| *Nle1* | Notchless homolog 1 (Drosophilia) | 143109 | NM_018096.2 | Hs00216436_m1 |
| *Oct4* | POU domain, class 5, transcription factor 1 | 157769 | NM_203289.2 | Hs01895061_u1 |
| *Pax6* | Paired box gene 6 (aniridia, keratitis) | 187128 | NM_000280.1 | Hs00240871_m1 |
| *Plxna2* | Plexin A2 | 221921 | NM_153014.1 | Hs00257877_m1 |
| *Scap2* | Src family associated phosphoprotein 2 | 226732 | NM_003930.3 | Hs00182698_m1 |
| *Sox9* | SRY (sex determining region Y)-box 9 (camponmelic dysplasia, autosomal sex-reversal) | 201106 | NM_000346.2 | Hs00165814_m1 |
| *Spry1* | Sprouty homolog 1, antagonist of FGF signalling (Drosophila) | 180417 | NM_005841.1 | Hs00398096_m1 |
| *Spry4* | Sprouty homolog 4 (Drosophila) | 130525 | NM_030964.2 | Hs00540086_m1 |
| *St7L* | suppression of tumorigenicity 7 like | 128921 | NM_198327.1 | Hs00373316_m1 |
| *Tdgf1* | Teratocarcinoma-derived growth factor 1 | 178078 | NM_003212.1 | Hs02339499_g1 |
| *Utf1* | Undifferentiated embryonic stem cell transcription factor 1 | 149070 | NM_003577.1 | Hs00747497_g1) |

### 2.6.3 Analysis of qPCR data

Two different methods of analysing data from qPCR experiments exist – absolute and relative quantitation (Schmittgen et al., 2008). Absolute quantitation determines the input copy number of the transcript of interest, usually by relating the PCR signal to a standard curve. The data are typically presented as copy number per cell. When precise quantities of an amplicon are required, for example, the calculation of viral load, absolute quantitation is used with the disadvantage being the added effort to generate standard curves. Relative quantitation describes the change in expression of the target gene relative to some reference group such as an untreated control or a sample at time zero in a time-course study. The $2^{-\Delta\Delta CT}$ or comparative $C_T$ method, a derivation of which is offered by Livak et al, 2001, is used to calculate relative

changes in gene expression determined from real-time quantitative PCR experiments (Livak et al., 2001). It has the advantages of being easy to use and the ability to present data as 'fold change' in expression. However, there are some disadvantages which include the assumption of PCR efficiency which must hold or the PCR must be further optimised. Relative quantitation was used with the $2^{-\Delta\Delta CT}$ or comparative $C_T$ method in this study.

Equations used in the comparative $C_T$ method:

- Fold change = $2^{-\Delta\Delta CT}$
- $2^{-\Delta\Delta CT}$ = [($C_T$ gene of interest – $C_T$ internal control) sample A – ($C_T$ gene of interest – $C_T$ internal control) sample B)]

Use of this equation can be used to compare the gene expression in two different samples such as sample A, undifferentiated cells and sample B, differentiated cells.

# CHAPTER 3 CELL CULTURE DESIGN AND OPTIMISATION

## 3.1 Introduction

With the concept of cancer stemness and the existence of a cancer stem cell as the progenitor cell of any given tumour being more widely accepted, the identification of a target/treatment that would potentially eliminate these cells thus eradicating cancer would constitute a major advance in the treatment of cancer and prevention of recurrence and metastasis. Teratocarcinoma or embryonal carcinoma cell lines are often referred to as the malignant version of embryonic stem cells and possess the stem cell properties of self renewal and capacity to differentiate into virtually all tissue types, i.e. pluripotency. The comparison of the embryonic stem cell and its malignant counterpart may provide invaluable insight into the characteristics and properties of the cancer stem cell that may ultimately be used in its destruction in the treatment of cancer.

The in vitro study of embryonic stem cells and teratocarcinoma stem cells has many differences and intricacies compared with other cell lines in culture. Teratocarcinoma and embryonic stem cell lines in both mouse and human can be grown on a mitotically inactive layer of mouse fibroblasts to maintain the cells in an undifferentiated state.

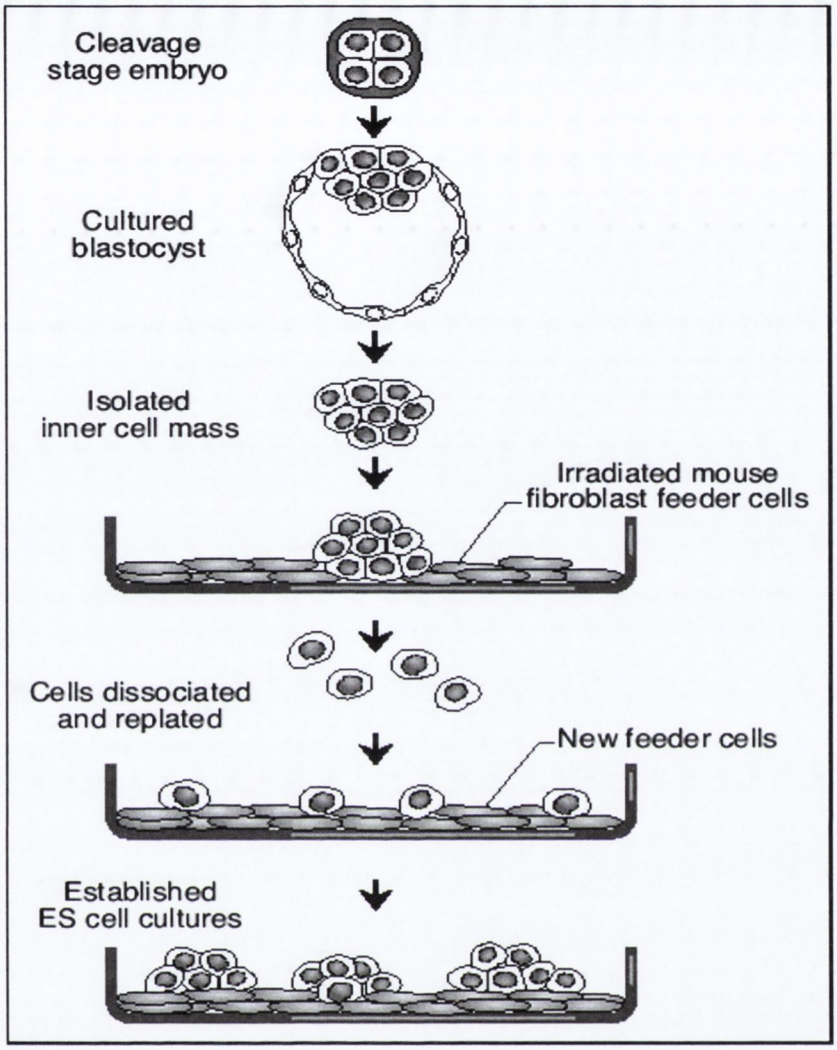



**Figure 3.1: This figure demonstrates the origin of ES cells and how they are subsequently cultured in vitro (Odorico et al., 2001).**

Growth factors released by the fibroblasts along with the presence of foetal bovine serum appear to be vital for the maintenance of the self-renewing state. Removal of the fibroblasts allows differentiation occur. However, not unexpectedly, the use of fibroblasts with these cell lines introduces a confounding factor in any further analysis of the cells of interest. This is particularly important in human cell culture populations where the use of human ES cells may be used for tissue regeneration and replacement purposes. The presence of mouse fibroblasts in such instances may be problematic. Ideally, feeder free conditions need to be in place thus preventing any source of contamination. Leukaemia inhibitory factor (LIF), added to the growth medium, has been used by many groups to try and maintain cells in the self-renewing

undifferentiated state with its absence in the cell growth medium allowing differentiation to occur spontaneously (Shen et al., 1992). In this study, attempts were made to grow cells using LIF in feeder free conditions and these are described. Human cell lines were grown without feeder layers using the differentiation protocol designed by Peter Andrews and his group.

A big step in establishing the optimal conditions is in determining if differentiation can be both prevented and encouraged.

### 3.2 Aims

- To determine optimal cell culture conditions for mouse and human cell lines.
    - Optimise mouse cell culture using feeder free conditions.
    - Optimise mouse cell culture in the presence of feeder cells, i.e. as per protocol from the ATCC.
    - Optimise human cell culture protocols.
- To determine and validate a suitable time course experiment prior to expression array analysis.
    - Determine optimal time point to study differentiation.
    - Successfully develop a time course experiment prior to expression array analysis.

### 3.3 Materials and methods

#### 3.3.1 Cell culture and morphology

Initially, all five cell lines were grown as per protocol outlined in Chapter 2. The protocol was adjusted to grow mouse cells in feeder free conditions by using LIF, instead of feeder cells for both SCC-PSA1 and Nulli-SCC1 cell lines. This was performed using the mouse cell lines only as protocols given for human cell lines did not require feeder cells or LIF.
Therefore mouse cell lines were grown in two ways:

a) With LIF in the absence of feeder cell layers.

b) Without LIF, in the presence of feeder cell layers.

**Table 3.1: Cell lines used in cell culture experiments.**

| *Cell lines* |
|---|
| **Mouse** |
| 1. SCC-PSA1, pluripotent teratocarcinoma cell line |
| 2. Nulli-SCC1, nullipotent teratocarcinoma cell line |
| 3. ES-E14TG2a, embryonic stem cell line. |
| **Human** |
| 4. Ntera-2 clone D1 – NT2/D1, human pluripotent teratocarcinoma cell line. |
| 5. 2102Ep, human nullipotent teratocarcinoma cell line |

Morphology was observed daily for all cell lines using a phase contrast microscope. Mouse cell lines were photographed at days 0, 3, 6, 9 and 21 in their differentiated and undifferentiated states with and without LIF and feeder cells. Human cell lines were photographed at similar intervals in their differentiated and undifferentiated states.

### 3.3.2 Time course experiments

The following time course experiments were performed using the differentiation and harvesting protocols as outlined in chapter 2:

Mouse time course experiments:

1. SCC-PSA1, pluripotent teratocarcinoma cell line:
    A. Day 0, 1, 2, 3 with and without LIF.
    B. Day 0, 3, 6, 9 with and without LIF.
    ***C. Day 0 and 3 with and without feeder cells (in triplicate).***

2. Nulli-SCC1, nullipotent teratocarcinoma cell line:
    A. Day 0, 1, 2, 3 with and without LIF.
    B. Day 0, 3, 6, 9 with and without LIF.
    ***C. Day 0 and 3 with and without feeder cells (in triplicate).***

3. ES-E14TG2a, embryonic stem cell line:
    ***A. Day 0 and 3 with and without feeder cells (in triplicate).***

Human time course experiments:

1. Ntera-2 clone D1 – NT2/D1, human pluripotent teratocarcinoma cell line.
    ***A. Day 0 and 3 with and without retinoic acid.***

2. 2102Ep, human nullipotent teratocarcinoma cell line.
    ***A. Day 0 and 3 with and without retinoic acid.***

(*Time courses in italics refer to final time course used in expression array analysis.)

### 3.3.3 PCR analysis

Reverse transcription PCR (RT-PCR) and subsequently quantitative real-time TaqMan® PCR (qRT-PCR) were performed on all mouse time course experiments within the SCC-PSA1 and Nulli-SCC1 cell line groups using the markers listed in Table 3.2. RT-PCR and quantitative real-time PCR were used to determine the differentiation status of cells at each time point in the experiments thus enabling determination of successful optimisation of cell culture protocols with and without the use of LIF and feeder cells.

**Table 3.2: Markers used to determine differentiation status in mouse time course experiments. (*The Nanog assay was not successfully optimised despite multiple attempts.)**

| Markers of self renewing state | Markers of differentiation |
|---|---|
| 1. Oct4 | 1. Afp (endoderm) |
| 2. Sox2 | 2. Ncam (ectoderm) |
| 3. Nanog * | 3. Vegfr2 (mesoderm) |

Marker status was not performed on ES-E14TG2a, mouse embryonic stem cell line Ntera-2 clone D1 – NT2/D1, human pluripotent teratocarcinoma cell line or 2102Ep, human nullipotent teratocarcinoma cell line. The information obtained from the extensive analysis of SCC-PSA1 and Nulli-SCC1, pluripotent and nullipotent mouse teratocarcinoma cell lines was deemed sufficient to only perform the final time course experiment on the remaining cell lines.

## 3.4 Results

### 3.4.1 Methods to monitor morphology

Methods to monitor morphology include:

#### 3.4.1.1 Observation of morphology by macro and microscopy

#### 3.4.1.2 Use of differentiation markers to determine differentiation status using

a. Reverse transcription PCR.
b. Quantitative realtime PCR.

### 3.4.1.1 Observation of morphology by macro and microscopy

Cell lines were grown as per protocols outlined in Chapter 2. Morphology was observed both macroscopically and microscopically. Microscopic observations were performed using phase contrast microscopy and observations were made daily. Photographs were taken at daily intervals for the first three days and every three days subsequently.

***Mouse embryonic stem cells, ES-E14TG2a***

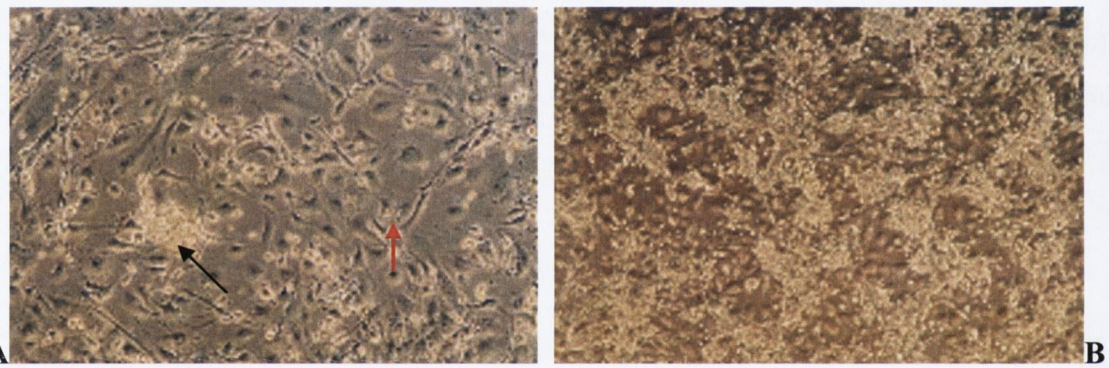



**Figure 3.2: mES cells: A) Day 1 post subculture on feeders (black arrow points to a cluster of mES cells, red arrow points to a feeder cell). Phase contrast, 10x. B) Day 3 post subculture on feeders. Phase contrast, 10x.**

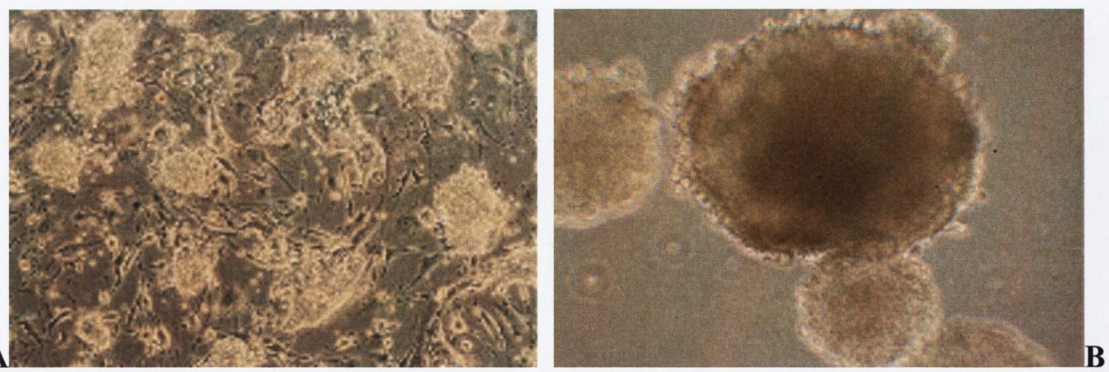



**Figure 3.3: mES cells: A) Day 3 differentiated. B) Day 9: Differentiated mES cells forming embryoid bodies in Petri dish without feeders.**

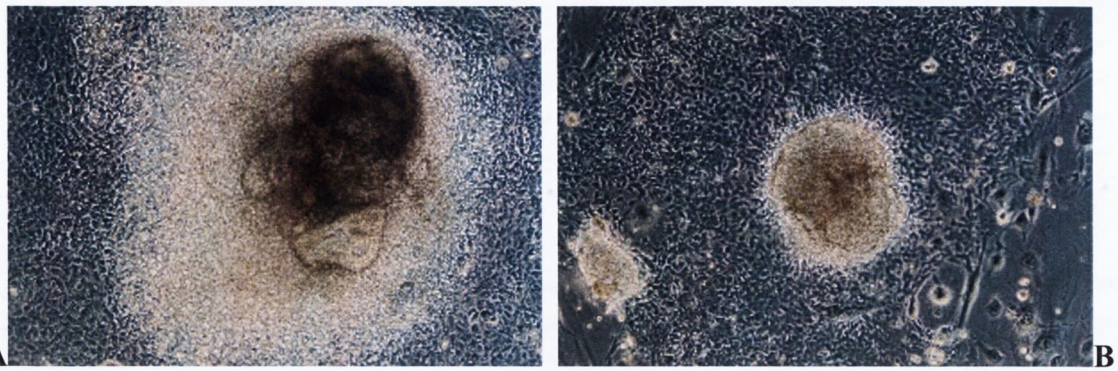



**Figure 3.4: mES cells: A) Day 25 Differentiated mES cells grown without feeders (pulsations noted in this cell group microscopically). B) Day 25 Differentiated mES cells.**

Several attempts were required to successfully grow and stock mES cells. Some additional measures were required to achieve optimal growth following a search of the literature and consultation with colleagues in a sister university. The medium was changed every day instead of every second day as with the other cell lines. Medium was changed 3 – 4 hours prior to passaging and harvesting. Additional pipetting was performed following trypsinisation to ensure a single cell suspension rather than cells clumping together. Careful observation of the cells and the addition of these extra measures allowed successful growth of this cell line. Cells were allowed to differentiate for up to 25 days. Development of cardiomyocytes occurred with pulsation of the cell groups observed. ES cells are capable of differentiating into haematopoietic progenitors, yolk sac, skeletal myocytes, smooth muscle cells, adipocytes, chondrocytes, endothelial cells, melanocytes, neurons, glia, pancreatic islet cells and many more.

*SCC-PSA1, pluripotent teratocarcinoma cell line*

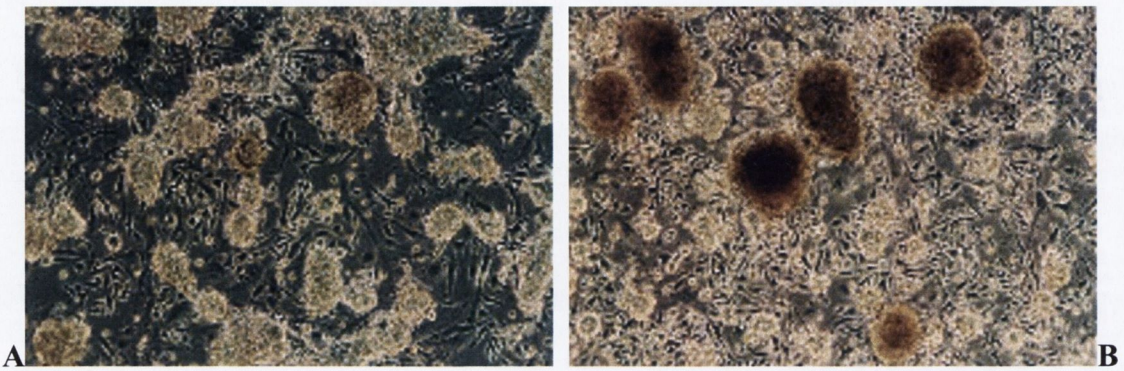



**Figure 3.5: SCC-PSA1: A) Day 0 on feeder layers, undifferentiated, 10x. B) Day 3 on feeder layer, undifferentiated, 10x.**

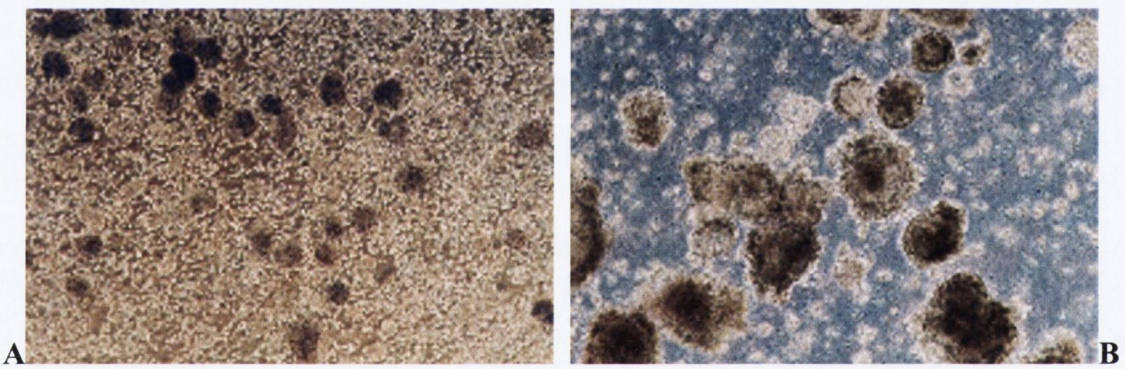



**Figure 3.6: SCC-PSA1: A) Day 3 without feeders, differentiated, 5x. B) Day 6 without feeders, 10x.**

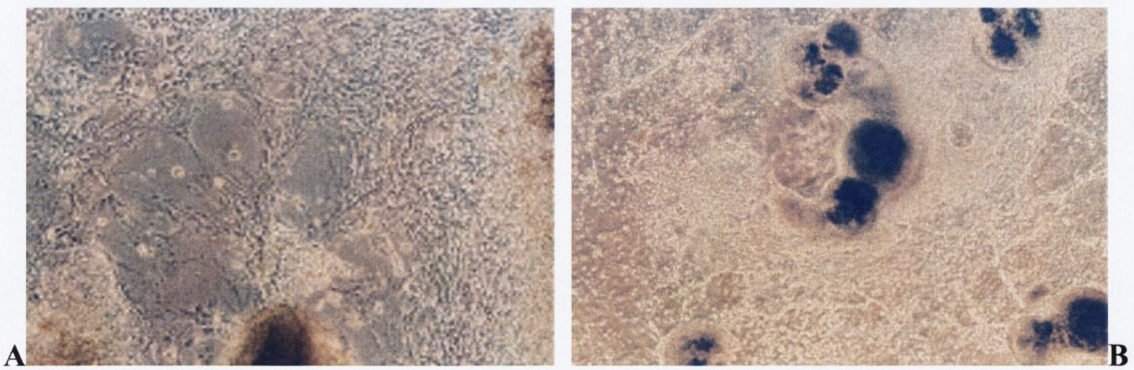



**Figure 3.7: SCC-PSA1: A and B) Day 25 differentiated.**

This cell line grew well in culture on feeder cell layers. Successful growth with the creation of stocks was reached quickly.

***Nulli-SCC1, nullipotent teratocarcinoma cell line***

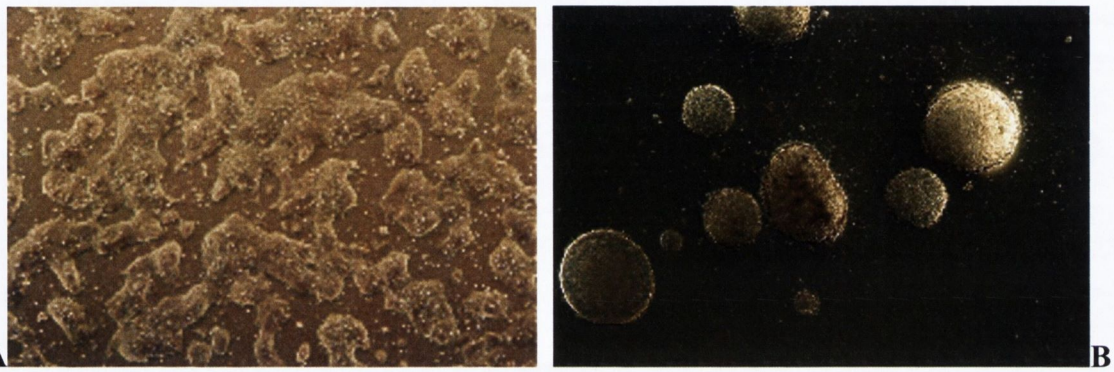



**Figure 3.8: A) Nulli-SCC1 day 3, 'undifferentiated', 5x. B) Nulli-SCC day 3, 'differentiated', 5x.**

While this cell line does not differentiate, the differentiation protocol used for the other two cell lines was applied as a control. Interestingly, differences were noticed in morphology as demonstrated in the photographs in Figure 3.8 when the undifferentiated and differentiated protocols were followed. Feeder cell layers were not used however in the undifferentiated protocol as they are not required for their growth. All flasks were gelatinised. Using the differentiation protocol growing cells in round 9cm bacteriological Petri dishes produced clear well defined round clusters of cells. These cells were also successfully grown with and without LIF as demonstrated in Figures 3.9 and 3.10.

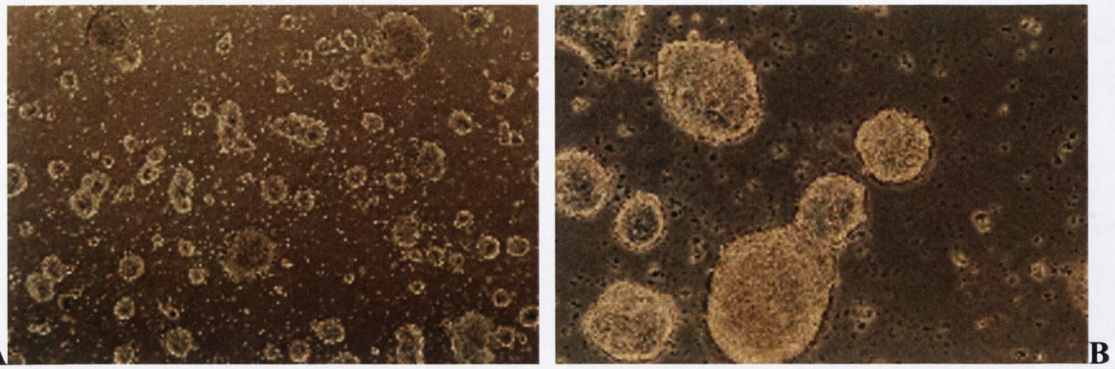



**Figure 3.9: A) Nulli-SCC1 day 3, with LIF, 5x. B) Nulli-SCC1 day 3, with LIF, 20x.**

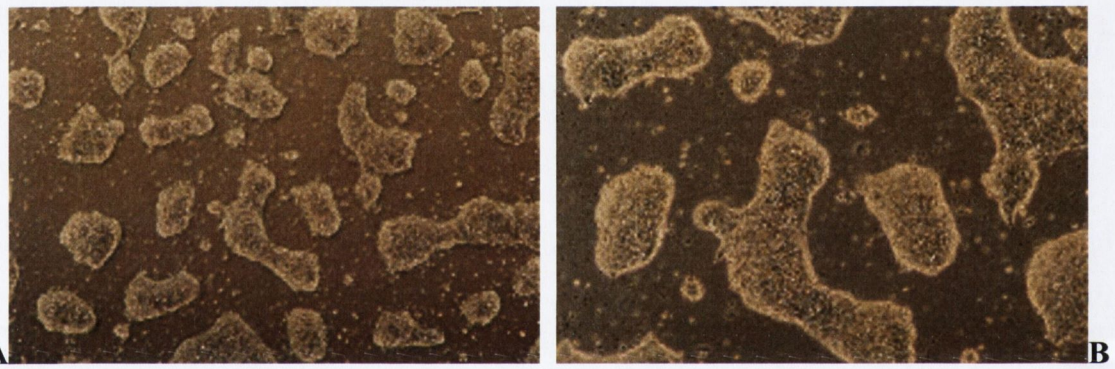



**Figure 3.10: A) Nulli-SCC1 day 3, without LIF, 5x. B) Nulli-SCC1 day 3 without LIF, 10x.**

Subtle morphological differences can be identified between those cells grown with and without LIF. In the cells grown with LIF, the cell groups are rounder and better defined with a definite outer rim around the cell groups (Figure 3.9). In the without LIF group (Figure 3.10), the cell groups have more irregular shapes and are not as well defined as those grown with LIF. The significance of these findings is uncertain and has not been previously described in the literature.

*Ntera-2 clone D1 – NT2/D1, human pluripotent teratocarcinoma cell line*

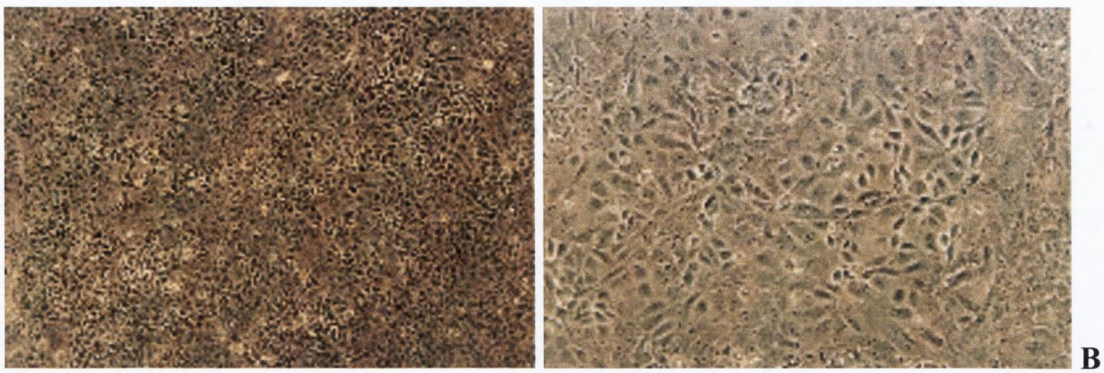



**Figure 3.11: A) Ntera2 cell line without retinoic acid at day 3, 10x. B) Ntera 2 cell line with retinoic acid at day 3, 20x.**

Almost no appreciable difference was detectable between Ntera2 cells grown with and without retinoic acid as is demonstrated in Figure 3.11. Cells grow in sheets covering the entire flask surface. As cell line protocols and time course experiments were optimised and determined for all the mouse cell lines, it was decided to follow the same time course for the human cell line experiments.

*2102Ep, human nullipotent teratocarcinoma cell line*


**Figure 3.12: A) 2102Ep cells at day 3, 5x. B) 2102Ep cells at day 7.**

Similar to the Ntera2 cell lines, there was very little morphological difference observed between cells grown with and without retinoic acid. This cell line grew easily with a slightly different morphology to Ntera2 cell lines. Cells grow in confluent sheets with smooth outlines to cell groups compared with sharper more well defined outlines in the Ntera 2 cell line. Interestingly, there is a similar morphology at low power to the mouse nullipotent teratocarcinoma cell line, Nulli-SCC1 (Figure 3.8A).

### 3.4.1.2 Use of differentiation markers to determine differentiation status

#### *a) Reverse transcription PCR*

Observation of morphological changes revealed changes indicative of commencement of the differentiation process. Maintaining the cells in a continuous undifferentiated state was more difficult to observe closely as the risk of an occasional cell or group of cells differentiating would not be clearly obvious on microscopic observation. To confirm that differentiation was occurring or not occurring as the case may be, expression of selected markers were used. Functional assays can also be performed such as colony forming assays and injection of cultured cells into blastocysts and recording their ability to produce chimeric mice (Palmqvist et al., 2005). A series of three markers expressed in differentiated and undifferentiated embryonic stem cells was used following an extensive search of the literature. RT-PCT was performed using Actin as a control, Oct4, Sox2, Nanog as markers of undifferentiation and Vegfr2, Afp and Ncam as markers of differentiation. Oct 4, Sox 2 and Nanog are widely used in stem cell literature as markers of differentiation. Initially, marker status was performed on SCC-PSA1 and Nulli-SCC1 cells as successful establishment of mES cells had not yet been achieved. This combination of markers of differentiation has been used previously by Abeyta et al, 2004, in hES experiments (Abeyta et al., 2004).

Oct4 is a transcription factor that belongs to the POU transcription factor family. Its expression is restricted to pluripotent cells of the inner cell mass and it is essential for the initial development of pluripotentiality in the inner cell mass (Pesce et al., 2001). Sox2 is another transcription factor that regulates pluripotency of mES cells and a reduction in its

expression stimulates differentiation (Masui et al., 2007). Nanog is yet another transcription factor and homeodomain protein also being a key regulator of pluripotentiality. While it is found in embryonic stem cells, it is largely absent from differentiated cells. Vegf or vascular endothelial growth factor has been implicated as the major growth factor for developing endothelial cells (Hirashima et al., 2003) and the Vegf system is thought to be a crucial regulatory system for angiogenesis (Hiratsuka et al., 2005). Vegfr2 is a Vegf receptor tyrosine kinase, a pivotal regulator of endothelial and haematopoietic development, and is expressed by mesodermal cells (Cortes et al., 1999; Hirashima et al., 2003; Izumi et al., 2007). Afp, alpha-foeto-protein, is often used as a marker of endodermal differentiation due to its expression by the visceral endoderm, yolk sac and foetal hepatocytes (Abe et al., 1996; Scohy et al., 2000; Kuai et al., 2003). Ncam, neural cell adhesion molecule 1, plays a role in controlling the proliferation of neural progenitor cells and directing their differentiation towards neurons or neuroectoderm and was found to be upregulated upon differentiation of mouse ES cells (Amoureux et al., 2000; Carpenter et al., 2001; Ahn et al., 2004; Capkovic et al., 2008).

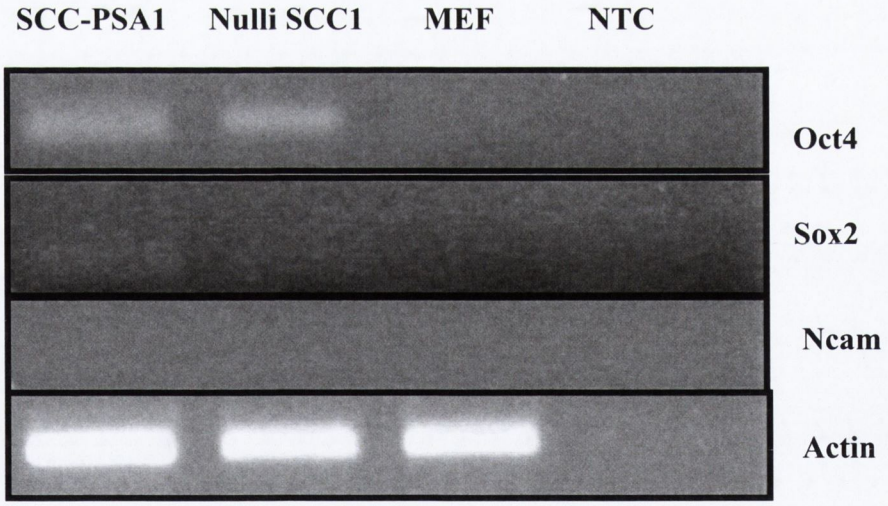



**Figure 3.13: Expression of markers of pluripotency.**

The panels represent RT-PCR analysis of the expression of markers of pluripotency (Oct4, Sox2) and differentiation (Ncam) in undifferentiated pluripotent teratocarcinoma cells (SCC-PSA1), nullipotent cells (Nulli-SCC1) and mitotically inactive feeders (MEF). The expression of Actin is included as an internal control. All samples were separated on a 2% agarose/EtBr

gel. As expected in undifferentiated cells, the differentiation marker, Ncam, is not expressed while there is expression of both Oct 4 and Sox 2, although weak in nature.

Ultimately, RT-PCR proved time consuming and difficult to accurately reproduce so we moved on to a more accurate and quantitative method using real-time PCR.

***b) Quantitative Real-time RT-PCR***

TaqMan® assays were purchased from Applied Biosystems off the shelf for rodent Gapdh, Oct4, Sox2, Afp, Vegfr2 and Ncam. An assay for Nanog had to be custom designed. Time point experiments on SCC-PSA1 and Nulli-SCC1 were performed as outlined in section 3.3. Again, experiments were confined to these two cell lines while mES cells were being grown and stocked. Once conditions were optimal for SCC-PSA1 cell line, they could be applied to the mES cell line.

Before doing the actual assays, the following items were considered:

1. Choosing a suitable internal control.
2. Ensuring no contaminating DNA remained in the RNA samples.
3. Assessing the efficiency of the TaqMan® assays.

- Choosing a suitable internal control:

As Gapdh and Actin are commonly used as internal controls, we started off with assessment of two rodent Gapdhs and β actin. It is important to ensure that there is no change in the expression of the internal control used over the course of the experiment and that is not affected by the given treatment.

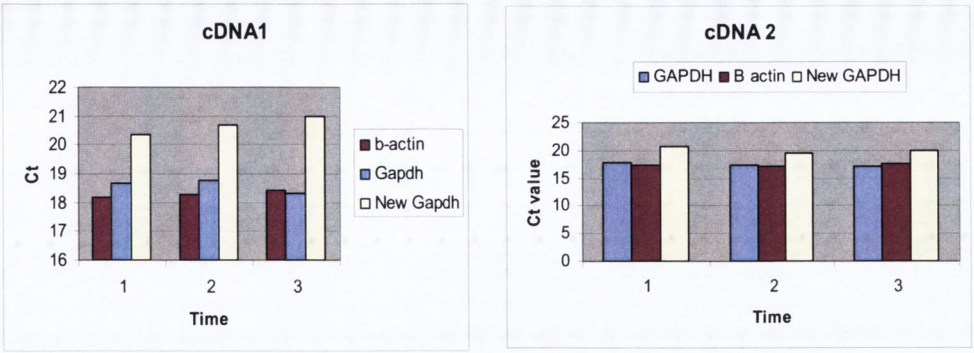



**Figure 3.14: This experiment uses cDNA from the SCC-PSA1 time courses B and C. cDNA1 = cDNA from SCC-PSA1 time course B, cDNA2 = cDNA from SCC-PSA1 time course C. 1 = 0hr, 2 = 72hr + LIF, 3 = 72hr – LIF.**

Based on this experiment, β-actin and both rodent Gapdhs did not change significantly over the 72hr time period illustrating that any of these assays could be used as an internal control. Gapdh was used in all subsequent experiments as it was readily available and extensively used within the laboratory.

- Ensuring no DNA contamination in RNA samples:

Reverse transcription (RT) + (positive) and RT – (negative) experiments were performed. cDNA was used as the test sample in the RT + samples and pure total RNA was used as the test sample in RT – experiments. All RNA samples from all the time course experiments were tested. Only a small number of samples had to undergo repeat DNase treatment and RT+/RT- experiments as a result of these experiments thus ensuring no DNA contamination within samples prior to testing samples further.

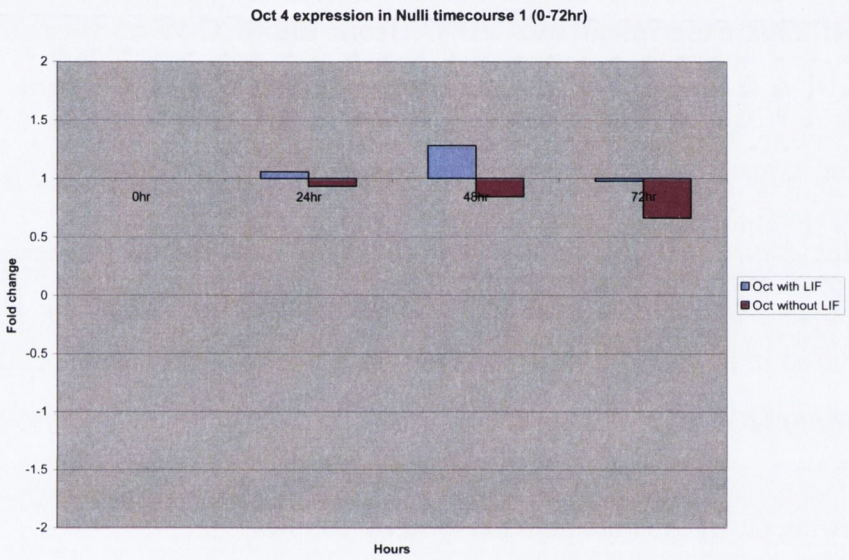
- Assessing the efficiency of the TaqMan® assays:

TaqMan® experiments to compare efficiency of assays with the rodent Gapdh assay were performed. To proceed, the efficiencies would have to be within 0.1 of each other. However, it was subsequently noted with off the shelf assays designed by AB that testing efficiency is not necessary as they have undergone extensive efficiency testing before development by AB and therefore it is not necessary to repeat this work. However, as Nanog had to be custom designed, optimisation and efficiency testing were undertaken and were successfully achieved.

Finally, it was possible to go ahead and test the time course experiments with the assays purchased and evaluate results.

***Nulli-SCC1 time course results***

**Oct4:**




A

**Figure 3.15: Oct 4 expression in Nulli time course A (0, 24, 48 and 72hrs). Blue = with LIF, purple = without LIF.**

This graph with fold change on the y axis and time on the x axis demonstrates that there is no significant change in Oct 4 expression in this cell line over 72 hours or in the presence of absence of LIF. This result would be as expected in a cell line known not to differentiate.

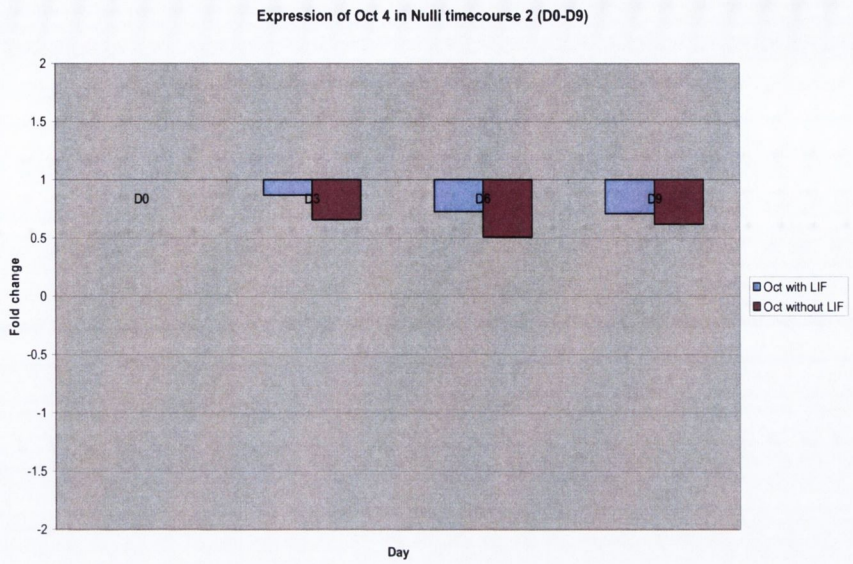



**Figure 3.16: Oct 4 expression in Nulli time course B (Day 0, 3, 6 and 9). Blue = with LIF, purple = without LIF.**

There is no significant change in Oct 4 expression over time or in the presence or absence of LIF. This is as expected and compares favourably with the 72 hr time course experiment.

**Sox2:**


**Figure 3.17: Sox2 expression in Nulli time course A (0-72hr) with (blue) and without (wine) LIF.**

This graph illustrates Sox2 expression over a 72hr time period in the presence and absence of LIF. The 72hr with LIF sample shows a fold change slightly greater than 2, the remaining samples show no significant changes. Fold changes over 2 are generally considered significant. As the change here is so minor, it was not considered significant and the expression of Sox2 in subsequent time courses would be observed carefully.

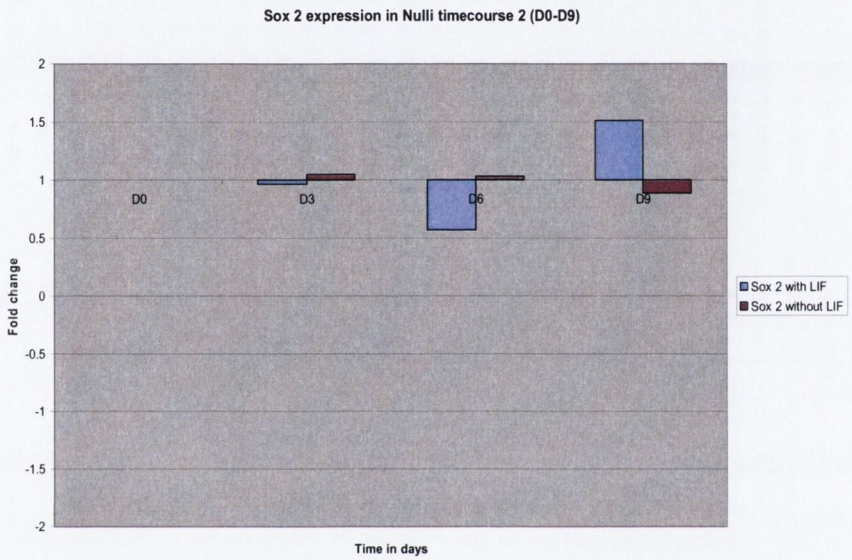



**Figure 3.18: Expression of Sox2 expression in Nulli time course 2 (D0, 3, 6 and 9).**

There is no significant change in Sox2 expression over time or in the presence or absence of LIF. In the 72 hr sample of the 72hr time course, there was a slight increase in Sox2 expression; however this change was not reproduced in this 9 day time course.

## Vegfr2:

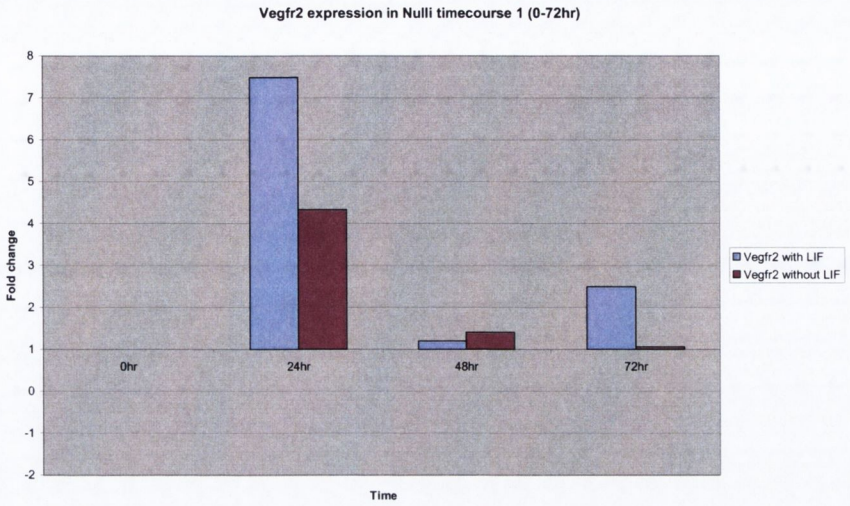



**Figure 3.19: Vegfr2 expression in Nulli time course A (0-72hr) with and without LIF.**

The first notable comment is that Vegfr2 appears to be expressed at all time points in undifferentiated teratocarcinoma cells unlike the other differentiation marker AFP which was not expressed. There appears to be a significant increase in expression at 24 hr with and without LIF with a slight significant change at 72 hr with LIF. Thus, overall, the results show an increase in Vegfr2 expression at 24 hr with a subsequent reduction in Vegfr2 expression over time. However, in this experiment, Vegfr2 was being expressed at cycle numbers of 35-39 with poor reproducibility between replicates particularly in the significant samples; therefore the significance of these results is uncertain.

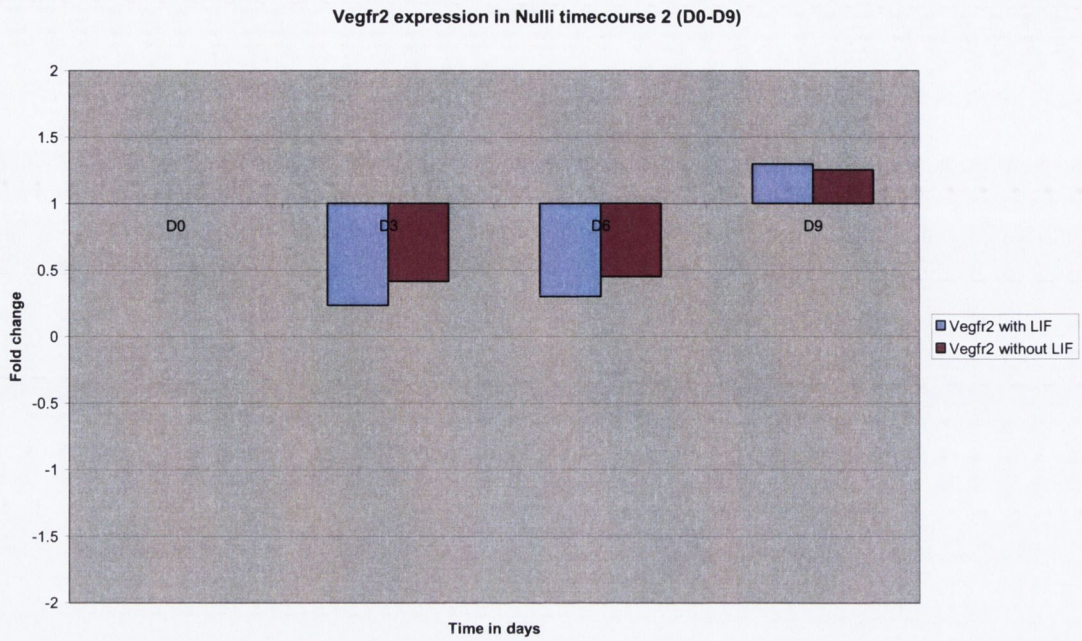



**Figure 3.20: Expression of Vegfr2 in Nulli time course B (D0-D9).**

Again, similar to time course A, the first point to note is that Vegfr2, a marker of differentiation is expressed at all time points. However, in this experiment, no significant changes in expression are noted. Therefore, it is likely that the results in time course A were due to the high cycle numbers obtained.

**Ncam and Afp:**

Both of these markers of differentiation were found not to be significantly expressed in Nulli-SCC1 cells. This is not an unexpected result given that Nulli-SCC1 cells are not capable of differentiation.

**Comments on Nulli-SCC1 data:**

The Oct4 and Sox2 data on this cell line are in keeping with what would be expected i.e. no significant change in expression over time given that no differentiation has occurred. Both time course experiments appear to correlate with each other with some minor differences. While there is no significant change in Vegfr2 expression following the differentiation protocol in nullipotent cells, it is interesting to note that there appears to be a baseline

expression of Vegfr2 in undifferentiated teratocarcinoma cells. The reason for this is not entirely clear but may reflect the possibility of cells being present with the capability for differentiation and thus possessing the Vegfr2 receptor. Afp was not found to be expressed by Nulli-SCC1 cells in these experiments (data not shown graphically). An occasional positive sample at cycle numbers in the region of 38 or 39 was detected but no significant changes were identified. Again, this absence of Afp is in keeping with the literature which reports Afp expression to be seen only upon differentiation in both ES and EC cells (Abe et al., 1996).

In summary, Nulli-SCC1 was cultured successfully as per protocol as well as successfully culturing it in the presence of LIF. As Nulli-SCC1 is a nullipotent cell line and thus incapable of differentiation, LIF would not be expected to have a significant effect on these cells. TaqMan® data using self renewal and differentiation markers are as reported and expected according to published literature. Distinct morphological changes were noted on microscopic observation when these cells were grown in the presence and absence of LIF. However, this morphological observation was not reflected by changes in the TaqMan® analysis of the expression of the differentiation and undifferentiation markers.

***SCC-PSA1 time course results***

**Oct4:**

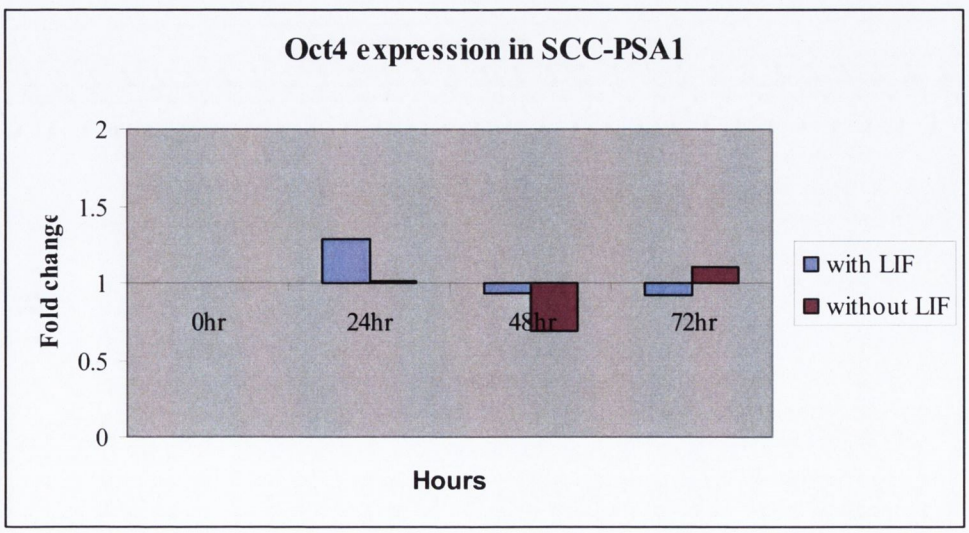



**Figure 3.21: Oct4 expression in SCC-PSA1 time course A (0, 24, 48 and 72 hours).**

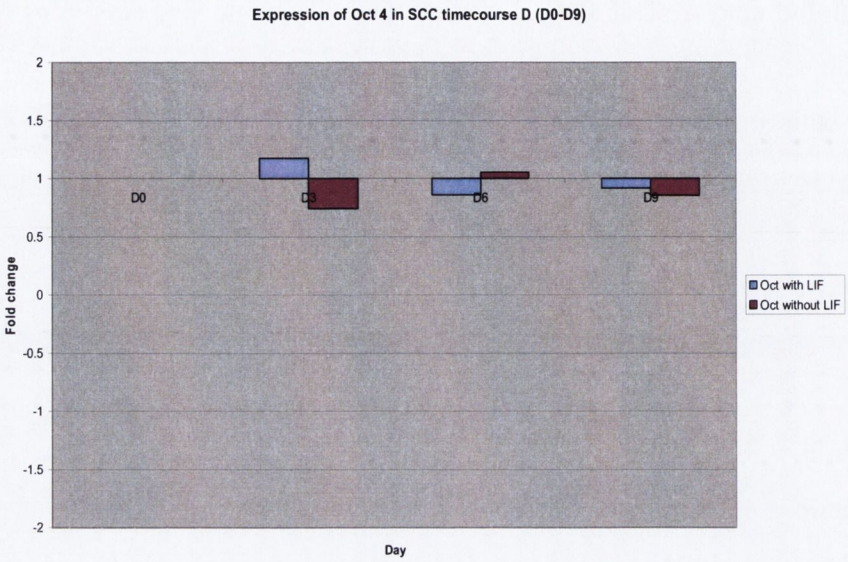



**Figure 3.22: Oct4 expression in SCC time course B (D0, 3, 6 and 9).**

## Comment on Oct4:

While Oct4 is a marker of the self renewing state, its expression might be expected to decrease over time. However, in both time courses illustrated with a time course extending to 9 days, no significant change in expression was seen at any stage. Palmqvist et al 2005 showed similar findings at 72 hours following removal of LIF with no significant decrease or increase in expression between 0 and 72 hours (Palmqvist et al., 2005).

**Sox2:**

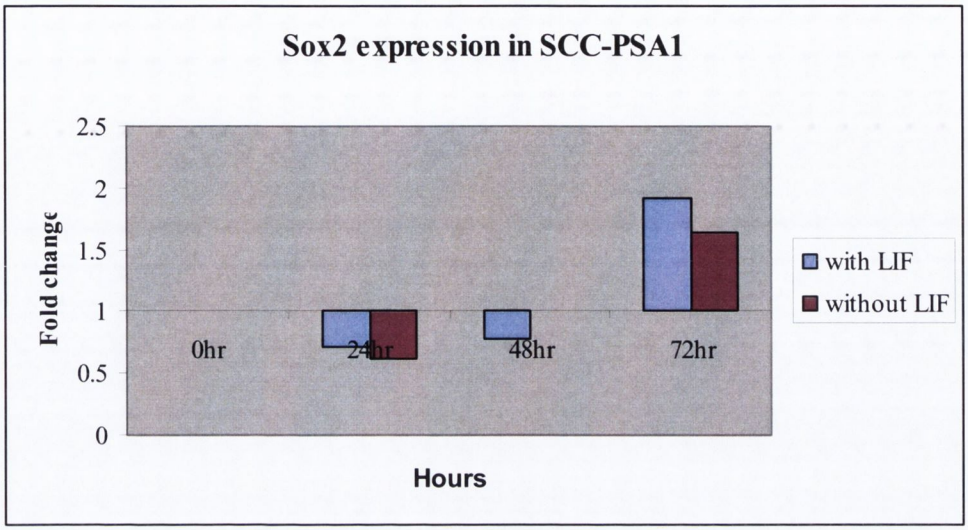



**Figure 3.23: Sox2 expression in SCC-PSA1 time course A (0, 24, 48 and 72 hours).**

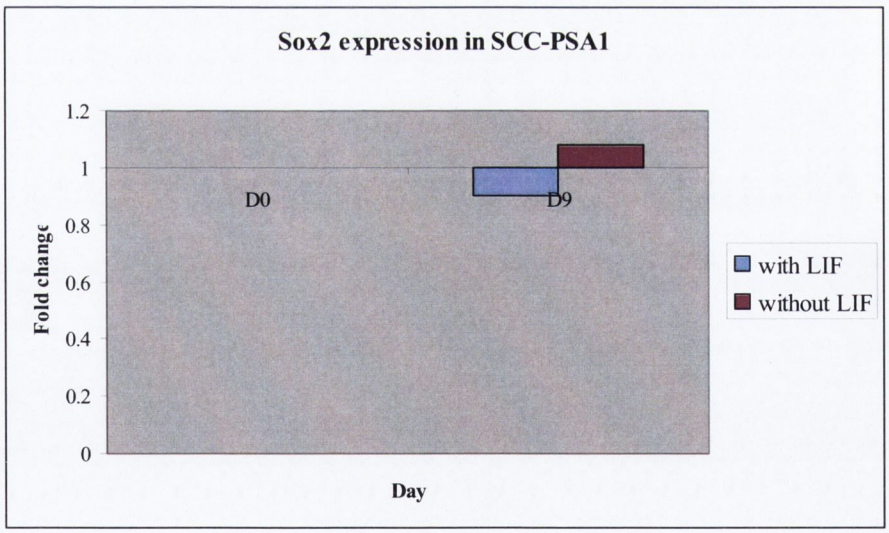



**Figure 3.24: Sox2 expression in SCC-PSA1 time course B adjusted (Day 0 and Day 9 only).**

**Comment on Sox2:**

Sox2 in time course A showed no significant change in expression at 24 or 48 hours compared with 0 hours. A slight increase in expression is seen with LIF at 72 hours with no significant change without LIF. Palmqvist et al 2005 found a significant decrease in Sox2 expression between 0 and 72 hours following the removal of LIF (Palmqvist et al., 2005). The significance of the marginally significant increase in Sox2 expression with LIF at 72 hours in Time course A is uncertain and was not repeated in Time course B (data not shown).

**Afp:**

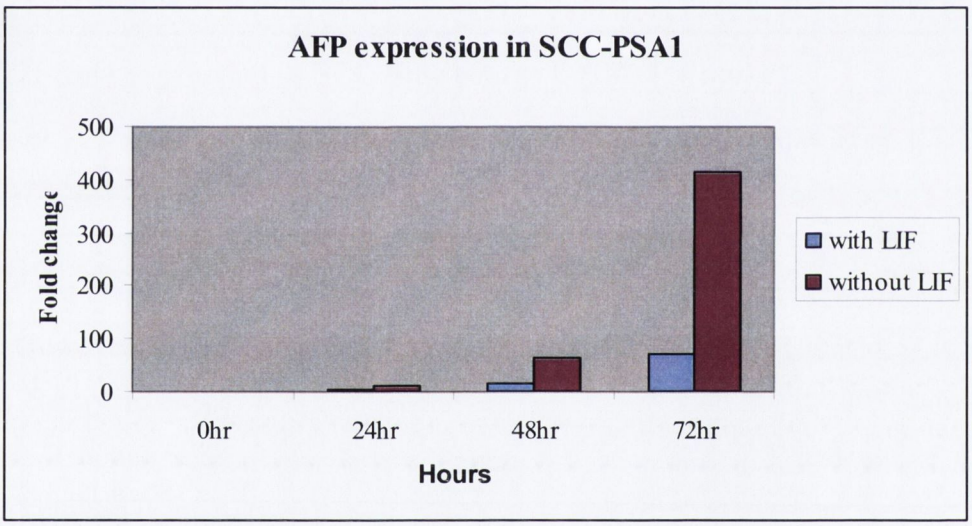



**Figure 3.25: Expression of Afp in SCC time course A (0, 24, 48 and 72 hours).**

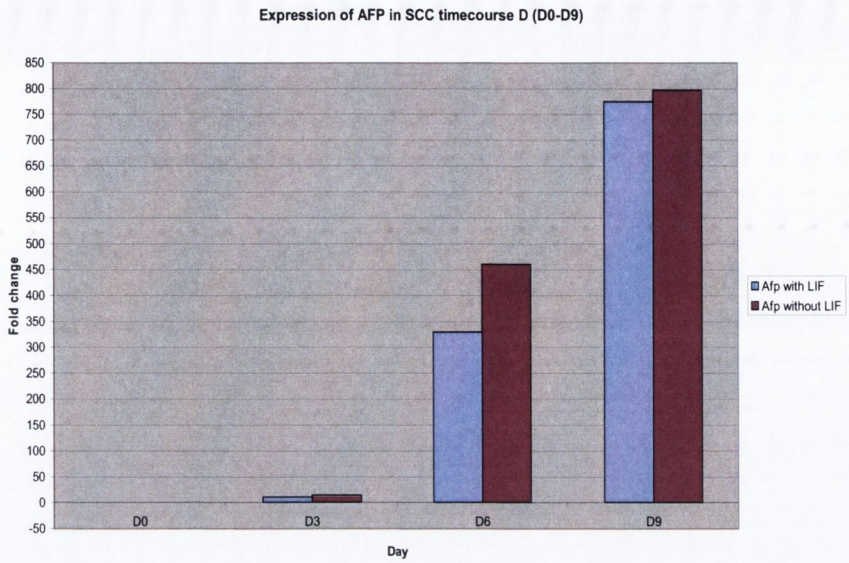



**Figure 3.26: Expression of Afp in SCC time course B (D0, 3, 6 and 9). Blue = with LIF, purple = without LIF.**

**Comment:**

The results with Afp are more dramatic. A highly significant increase in Afp expression is seen at all time points in each time course experiment. According to the literature, an increase in Afp upon differentiation would be expected, and is usually seen at approximately 5 days plus (Abe et al., 1996; Chinzei et al., 2002). LIF did not appear to have any effect on Afp expression with increases in expression seen with both LIF and without LIF.

**Ncam:**

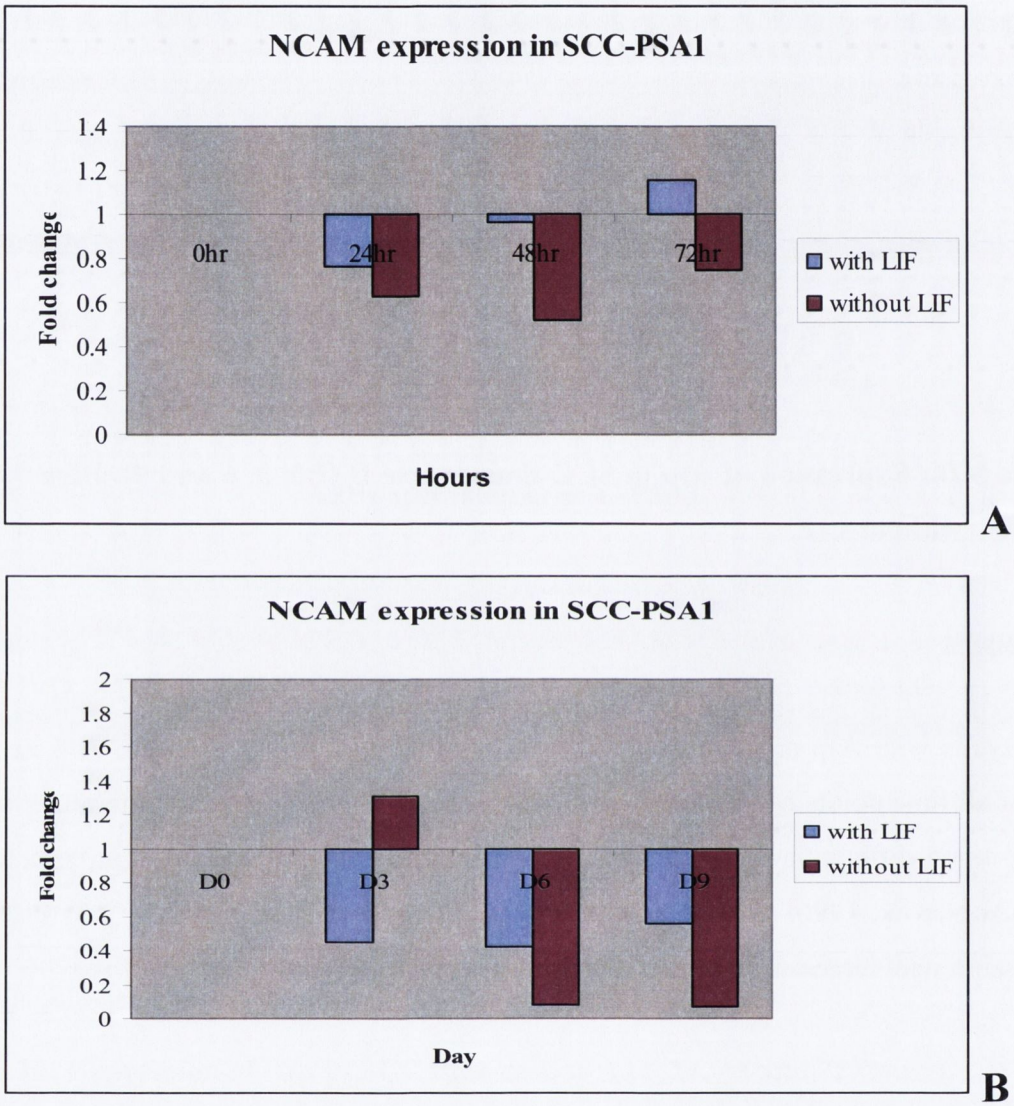
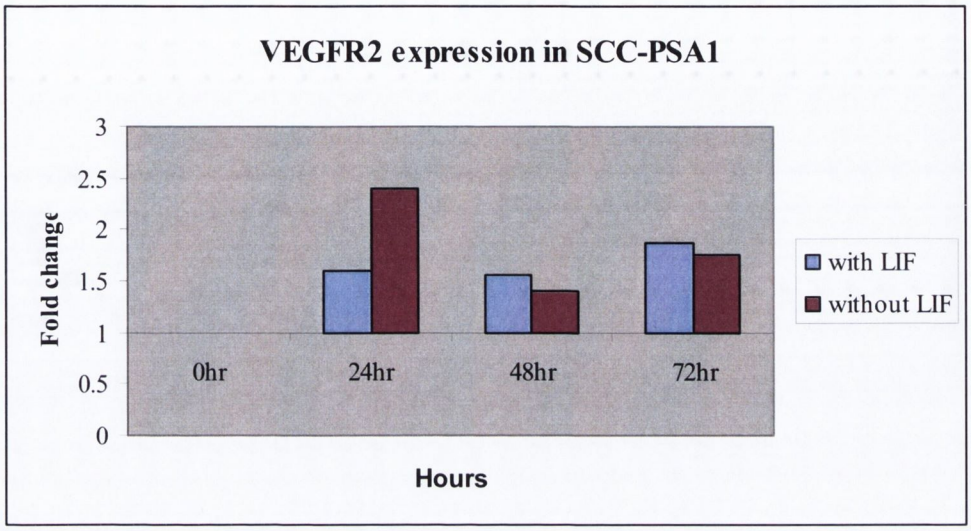


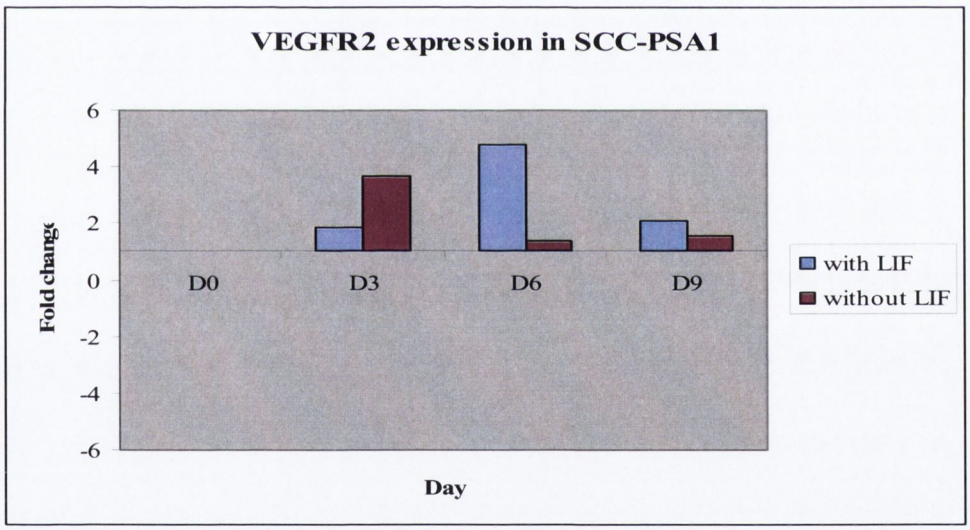

**Figure 3.27: Ncam expression in A) SCC-PSA1 time course A (0, 24, 48 and 72 hours) and in B) SCC-PSA1 time course B (Day 0, 3, 6 and 9).**

Both time courses show a similar result with no significant change in expression of Ncam over time with or without LIF.

**Vegfr2:**




A




B

**Figure 3.28: Vegfr2 expression in A) SCC-PSA1 time course A (0, 24, 48 and 72 hours) and B) SCC-PSA1 time course B (Day 0, 3, 6 and 9).**

**Comment:**

Vegfr2 expression at 24 hours is almost significant with a fold change value exceeding 2 slightly; however no further increase in expression is seen at 48 or 72 hours. An increase in expression is seen at day 3 without LIF and at day 6 with LIF with expression returning to

below 2 at day 9. The significance of these results is uncertain. Without uniform results between experiments with and without LIF, the use of LIF in these experiments is questionable.

***Comparison of SCC-PSA1 and Nulli-SCC1 time course experiments:***

An interesting observation was that SCC-PSA1 cells appeared to have a lower baseline expression of Oct4 and Sox2 compared with the Nulli-SCC1 cells in these experiments. In all experiments, the same amount of RNA was used in all RT reactions. When the TaqMan® experiment was performed with the Nulli cell line, Oct 4 was being expressed at only 13 cycles, a lower cycle number than with SCC cell line samples, and therefore a diluted sample of cDNA was used, either 1:10 or 1:100. This was not the case with SCC in which undiluted cDNA samples were used for all experiments.

**Table 3.3: Delta CT values for Oct4 and Sox2 in Nulli and SCC cell lines at day 0 and day 9, + and – LIF (normalising to internal control, rodent Gapdh).**

| | *Oct4* | | *Sox2* | |
|---|---|---|---|---|
| | **Nulli** | **SCC** | **Nulli** | **SCC** |
| ***D0*** | 3.2838 | -0.6379 | 2.8264 | -1.0939 |
| ***D9 +LIF*** | 3.7462 | -0.4638 | 3.0635 | -0.9406 |
| ***D9 -LIF*** | 4.02 | -0.4466 | 3.1748 | -1.206 |

The table above shows the delta CT values for both Oct 4 and Sox 2 (markers of pluripotency) in Nulli-SCC1 and SCC-PSA1 cells. This shows the difference in baseline expression of these markers between the two cell lines as well as showing that no major changes are occurring over time either in the presence or absence of LIF. The explanation for these findings may be that LIF is having no effect and therefore no change in marker expression in noted. An alternative explanation might be that the nullipotent cell line requires increased expression levels of Oct4 and Sox2 to maintain a constant state of undifferentiation.

### 3.4.2 Decision on use of LIF versus use of feeder cells only

Within this series of experiments comparing marker expression at a series of time points from day 0 up to day 9, no significant increases in expression were seen with Vegfr2 and Ncam and no significant decreases in expression were seen with Oct4 and Sox2. The Nanog assay was unsuccessful despite numerous attempts with both RT-PCR and TaqMan® PCR. The only marker showing any significant change in expression over time was Afp, showing a marked increase in expression over time in SCC-PSA1 cell lines even at 3 days with and without LIF. Afp was not found to be expressed by Nulli-SCC1 cells.

Explanations for these results:

- The addition of LIF was having no effect on the cells.
- Early time points of differentiation selected, therefore significant changes may not have occurred.
- Had cells begun to differentiate during or prior to the experiments thus rendering results difficult to interpret?

However, all experiments illustrated showed no difference in expression of the selected markers between the individual time courses with or without LIF, regardless of the level of expression of the given markers. For example, a significant increase in Afp expression over time was identified in SCC-PSA1 cells and this was seen in both cells grown with and without LIF. In summary, similar results were obtained from each time course experiment performed. As the differentiation markers appeared to behave as expected and the repeated absence of Afp expression in Nulli-SCC1 cells, the possibility that the cells had already begun to differentiate was unlikely.

In conclusion, repeated analysis of markers of differentiation and undifferentiation in the presence and absence of LIF showed no significant differences in marker expression in multiple time course experiments and differentiation was demonstrated to be occurring in both the absence and presence of LIF. Optimisation of cell culture conditions in the presence of LIF was not successfully achieved and it was decided to recommence experiments using the

recommended protocol using feeder cell layers (i.e. the supplier's protocol) with careful observation of the cell morphology at all stages.

### 3.5 Discussion

Mouse ES cells are known to require more attention and consideration to be cultured in vitro. Continuous careful observation is vital to ensure that cells do not spontaneously differentiate and are maintained in the undifferentiated state. They grow in clusters as attached rounded masses thus it can be difficult to identify single cells (Koestenbauer et al., 2006). The essential function of the fibroblast feeder layer is to produce LIF thus enabling the ES cells to self-renew by activating the transcription factor STAT3 (Ying et al., 2003). Addition of other cytokines to cell medium results in more directed cell differentiation, for example BMP (bone morphogenetic proteins) promote epithelial differentiation and prevent neural differentiation (Ying et al., 2003). Addition of retinoic acid can direct cells towards neural differentiation. LIF null fibroblasts have been shown to be deficient at supporting self-renewal (Stewart et al 1992). A number of authors have replaced the requirement for feeders with the addition of LIF to cell culture media (Smith et al., 1988; Williams et al., 1988). Despite numerous attempts in this study, replacing feeder cell layers with addition of LIF to cell culture media was unsuccessful and the time course experiments performed for gene expression microarray analysis used fibroblast feeder layers.

Heo et al (2005) studied the transcriptome profiles of ES cells at various points following spontaneous differentiation, namely day 1, 7, 14, 21 and 28 days (Heo et al., 2005). The most dramatic changes were seen in the first two weeks, i.e. between 1, 7 and 14 days with close similarities between days 14, 21 and 28 thus suggesting that the process becomes stable after the second week. They found a large number of differentially regulated genes in the first week (1440 genes) with less following the second week (810 genes). The authors also noted a higher proportion of downregulated genes than that of the upregulated genes indicating that repression rather than activation of gene expression might be a dominant mechanism during spontaneous differentiation of ES cells. This finding was consistent with the findings in all comparisons performed in subsequent chapters of this thesis where the proportion of downregulated genes was always significantly higher than upregulated genes.

Palmqvist et al 2005 also studied the transcriptome profiles of ES cells at 0hrs and at subsequent time points following the removal of LIF, namely 18 and 72 hours (Palmqvist et al., 2005). They found no significant change in expression of Oct4 over time using both quantitative PCR and protein expression methods but found Sox2 and Nanog to significantly decrease in expression at 72 hours. These results are not unlike those obtained for EC cells in this chapter where no significant change in expression over 3 days was demonstrated for Oct4 and Sox2. For their differentiation markers, they found Brachyury (a mesodermal marker) and Nestin (ectodermal marker) to be the most useful with a 15 and 3 fold increase respectively between 18 and 72 hours.

Despite numerous repeated attempts, assessment of Nanog expression in this series of cell line time course experiments was not attainable. The assay obtained from Applied Biosystems for this experiment was custom designed and thus required optimisation and testing of its efficiency prior to use. Thus optimisation and testing were undertaken but reproducible results were not obtained on multiple cell line samples following an array of experiments using both RT-PCR and realtime RT-PCR. Nanog is a transcription factor that maintains mouse ES cells in the pluripotent state independently of LIF and which upon differentiation is expected to decrease in expression (Shi et al., 2006). It has also been shown that Nanog is regulated by an adjacent pair of highly conserved Octamer- and Sox-binding sites through an interaction between Oct4 and Sox2 (Kuroda et al., 2005).

Afp is a marker of endodermal differentiation, specifically hepatic differentiation and is active during embryonic development and suppressed after birth. Abe et al, 1996, found that during EB (embryoid body) formation, a signal for Afp was first seen at day 5 (Abe et al., 1996). Afp has been found to be expressed in EC cells following three days of differentiation (Soprano et al., 1988). Looking at the TaqMan® PCR experiments performed in this chapter, Afp was significantly over expressed following three days of differentiation in mouse EC cells supporting findings found in other studies. These results together with Abe et al finding AfpP expression in mES cells indicates that Afp may be one of the most useful markers of differentiation in mES and EC cell lines as well as being useful at early differentiation time points. Nullipotent EC cells, a cell line not capable of differentiation, did not show any expression of Afp confirming its use as a marker of differentiation in ES and EC cells. Afp expression also increased despite the presence of LIF in the cell culture medium which should

prevent differentiation. It was one of the major factors in the decision not to proceed with the use of feeder free conditions.

The other differentiation markers used in this chapter, Ncam and Vegfr2, did not show as dramatic changes as Afp. Indeed with Ncam, no significant changes in expression were observed even up to day 9 of differentiation. Vefgr2 showed some small changes but interestingly the highest level of Vegfr2 expression was found at day 6 differentiation in the presence of LIF. This result added further support to the decision not to use feeder free conditions in subsequent experiments. Vegfr2 has been described as a marker of mesodermal differentiation, however, its level of expression over time is less well documented as it may be the presence of its ligands that determine how successful differentiation into haemopoietic and vascular elements will ultimately be. Eichmann et al 1997 found that while Vegfr2 was a marker of mesodermal cells, they noted no significant increase in expression of VegfrR2 over time and attributed this finding to the dependence on its ligands such as Vegf, Vegf-C to F (Eichmann et al., 1997). This may be an explanation for the lack of change in expression of Vegfr2 over time in these experiments. Interestingly, Vegfr2 expression was noted in the nullipotent EC cells but with no change over time demonstrated. In conclusion, Vegfr2 is not the most useful marker of differentiation and possibly only indicates the presence of cells with potential to differentiate into vascular and haemopoietic elements in the presence of Vegf ligands.

Ncam was also found to show no significant difference in expression between EC cells at day 0 and day 3 following differentiation and even up to day 9. Ahn et al 2004 examined Ncam expression following directed differentiation of mES cells to cells of neural origin and expression of Ncam was first documented at 6 days following differentiation (Ahn et al., 2004). While Ncam is a marker of differentiation and in particular a marker of ectoderm, it may be that Ncam is not expressed as early as day 3 following differentiation or even day 9 in the absence of directed differentiation thus explaining the lack of a change in expression over time in this chapter's RT-PCR experiments. Ncam was not detected in undifferentiated pluripotent cells as shown in Figure 3.13 nor was it detected in nullipotent EC cells.

Overall, the cell culture model provided a reproducible 3 day time course of differentiation using a feeder cell layer to prevent differentiation as per supplier's protocol. Feeder free

conditions as produced in other studies were not found to be consistently reproducible. Afp proved the most useful marker of differentiation showing a significant increase in expression even at 3 days differentiation in pluripotent EC cells. The early time point of 3 days proved too early to show significant changes in other markers such as Oct2, Sox2, Ncam and Vegfr2.

### 3.6 Conclusion

The initial aim of this chapter was to achieve optimal cell culture conditions for all the mouse and human cell lines in Table 3.1. Mouse pluripotent and nullipotent cells were grown with and without LIF to observe its effect on cell growth and differentiation and to determine if feeder free conditions could be achieved. Following observation of cell morphology and examination of marker expression using both RT-PCR and TaqMan® PCR, it was demonstrated that feeder free conditions could not be achieved successfully and reproduced on a regular basis. Therefore, fibroblast feeder layers were used as per protocol from the supplier's for mES and SCC-PSA1 cells for the final expression array time course experiment. Human pluripotent and nullipotent cell lines were grown as per protocol with and without retinoic acid for differentiation for the final expression array time course experiment. In summary, optimal cell culture conditions were obtained for mouse and human cell lines according to the ATCC protocols and protocols obtained by Professor Peter Andrews, Sheffield, UK as outlined in chapter 2. Feeder free conditions for mouse cell lines were not achieved.

Given that the majority of gene changes appear to occur during the first two weeks of differentiation, the aim was to focus on this time period and in particular the earliest point at which it has been suggested that changes are occurring, i.e. 3 days following spontaneous differentiation. ***Observation of morphology and RT-PCR experiments reported here show that changes are occurring, subtle but present, at three days so we proceeded to perform gene expression array analysis of these samples from mES, SCC-PSA1 and Nulli-SCC1 cell lines as well as performing a similar experiment on human cells using Ntera2 and 2102Ep cell lines.*** Due to ethical reasons, experimentation on human ES cells could not be performed and data from Ntera2 and 2102Ep cells was compared with that in the current literature.

# CHAPTER 4 TRANSCRIPTOME PROFILE OF CANCER STEMNESS IN A MOUSE MODEL

## 4.1 Summary

Striking similarities between stem cells and cancer cells have led to the concept of the existence of a cancer stem cell, a concept that has since been documented in many tumours including breast, brain and prostate tumours. Teratocarcinomas are malignant tumours occurring predominantly in the testes composed of undifferentiated stem cells and mature tissues. In this thesis, cancer stemness was studied using the teratocarcinoma model of tumourigenesis. The gene expression profile of murine embryonic stem cell lines was compared with its malignant counterpart, murine teratocarcinoma cell lines. Validation was performed using real-time quantitative PCR. A list of 1,170 differentially expressed genes was obtained. Significant pathways involved in cancer stemness included oxidative stress and angiogenesis. Transcription factors and extracellular matrix molecules appeared prominently. Novel molecules have been highlighted including decorin, an extracellular matrix protein, which may provide opportunities for the investigation of innovative strategies in the future treatment of cancer.

## 4.2 Introduction

The stem cell theory of cancer is a relatively old hypothesis but has been neglected by the prevailing paradigm in the cancer field (Trosko et al., 2005). The concept of the existence of a 'malignant stem cell' as the cell of origin of tumours was initially proposed by Pierce as far back as 1974 (Pierce, 1974). Due to their longevity and specific self-renewing properties, it is believed they have a greater propensity to accumulate carcinogenic mutations compared with short-lived, differentiated cells and thus are an ideal target of the carcinogenic process (Gudjonsson et al., 2005). Some of the first evidence of the existence of cancer stem cells came from the haematological malignancies where only a small subset of cancer cells were shown to be capable of forming new tumours (Bonnet et al., 1997). Their existence has since been demonstrated in prostate cancer (Collins et al., 2005; Collins et al., 2006), breast cancer (Al-Hajj et al., 2003), brain tumours (Singh et al., 2003), gastric cancer (Radtke et al., 2005), malignant melanoma and osteosarcoma (Gibbs et al., 2005). The similarities between stem cells and cancer cells are striking and have been extensively documented in the literature

(Reya et al., 2001). The defining features of a stem cell are its inherent abilities to both self renew and differentiate, features also present in some cancer cells albeit in a less controlled manner. Thus applying our knowledge of the principles of stem cell biology could lead to a more extensive understanding of the regulation of cancer cell proliferation and progression and produce new targets for more effective cancer therapies.

Teratomas (benign) and teratocarcinomas (malignant) are germ cell tumours found in the gonads and occasionally in extragonadal sites along the midline. Teratocarcinomas are composed of embryonal carcinoma cells as well as areas of more mature differentiated tissue including cartilage and intestinal glands. Embryonal carcinoma (EC) cells represent the stem cells of teratocarcinoma (Astigiano et al., 2005) and have also been termed the malignant equivalent of embryonic stem cells (Andrews, 2002). As many of the molecular pathways underlying tumourigenesis are also involved in normal embryogenesis, the study of EC cells provides an excellent model to elucidate the mechanisms involved in tumour growth.

This chapter aims to provide an insight into cancer stem cell biology through the study of teratoma tumourigenesis in a mouse model by comparison of normal embryonic stem cells and their malignant counterparts, teratocarcinoma cells. Identification of the fundamental differences between normal stem cells and their malignant counterparts should provide important information towards the development of successful cancer therapies. The specific targeting of cancer stem cells and their complete elimination is now regarded as essential to improve prognosis and allow recurrence-free survival.

### 4.3 Aims

The aim of this chapter is to provide a unique insight into cancer stem cell biology through the study of teratoma tumourigenesis in a mouse model by comparison of normal mouse embryonic stem cells and their malignant counterparts, teratocarcinoma cells. This analysis should produce a novel transcriptome profile of cancer stemness thus highlighting those genes unique to the cancer stem cell that may aid in more successful future cancer treatments.

## 4.4 Materials and Methods

### 4.4.1 Experimental sample collection

The following three cell lines; murine embryonic stem cell line, ES-E14TG2a, pluripotent teratocarcinoma cell line, SCC-PSA1 and the nullipotent teratocarcinoma cell line, Nulli-SCC1 were cultured as per protocol (see chapter 2, sections 2.1.1, 2.1.2 and 2.1.3) and harvested from two consecutive passages in triplicate.

Samples were labelled as follows:

1. ***E = ES-E14TG2a (murine embryonic stem cell line)***
    a. E1 – E3: Samples in triplicate from 1st passage.
    b. E4 – E6: Samples in triplicate from 2nd passage.
2. ***S = SCC-PSA1 (pluripotent teratocarcinoma cell line)***
    a. S1 – S3: Samples in triplicate from 1st passage.
    b. S4 – S6: Samples in triplicate from 2nd passage.
3. ***C = Nulli-SCC1 (nullipotent teratocarcinoma cell line)***
    a. C1 – C3: Samples in triplicate from 1st passage.
    b. C4 – C6: Samples in triplicate from 2nd passage.

### 4.4.2 RNA extraction and in vitro transcription (IVT)

Total RNA was extracted from all 18 samples identified as described in section 4.4.1 and the integrity and quantity of extracted RNA was verified as outlined in Chapter 2, section 2.3.4. The mRNA fraction was then converted into labelled cRNA using the Applied Biosystems Chemiluminescent RT-IVT Labelling Kit, Version 2.0. This process involves three main steps:

1) Reverse transcription of mRNA to cDNA.
2) Synthesis of the second DNA strand to form double-stranded cDNA.
3) Reverse transcription, labelling and amplification of cDNA to labelled cRNA.

The last step utilises a T7 (oligo dT) primer that binds to the polyA tail of mRNA / cDNA derived from mRNA. Therefore only the mRNA fraction of the sample is amplified. The dUTP nucleotides within the reaction mixture are labelled with digoxigenin. Multiple transcription rounds result in the production of DIG-labelled cRNA (see Chapter 2, section 2.5.2 for more details). RT-IVT reactions typically yield 140 – 170µg labelled cRNA from a starting sample of 2µg of total RNA. Between 20ng and 2µg of total RNA and up to 10 µl of sample may be used for input to the RT-IVT reaction. Generally, 2µg of total RNA was used for each sample.

### 4.4.3 Chemiluminescence reaction

The Applied Biosystems Expression Array System uses chemiluminescence to detect hybridisation events; chemiluminescence offers enhanced detection sensitivity compared with other microarray detection technologies, such as fluorescence, due to very low read noise and low background chemiluminescence. The chemiluminescence detection method is outlined in detail in Chapter 2, section 2.5.3. In summary, the process involves the following steps (Figure 4.1):

1) Hybridisation of fragmented DIG-labelled cRNA (from the RT-IVT reaction) to probes on the microarray by incubating for 16 hours at 55°C.
2) Washing microarrays to remove unhybridised DIG-labelled molecules.
3) Addition of alkaline phosphatase-DIG antibody conjugate to the microarray. DIG antibody (and the attached alkaline phosphatase) binds to DIG-labelled cRNA.
4) Addition of substrate and chemiluminescent enhancer. Alkaline phosphatase hydrolyses the chemiluminescent substrate and emits light at a wavelength of ~458nm. The enhancer strengthens the signal which is proportional to the amount of target RNA.
5) Capturing of chemiluminescent images and analysis on the 1700 analyser.

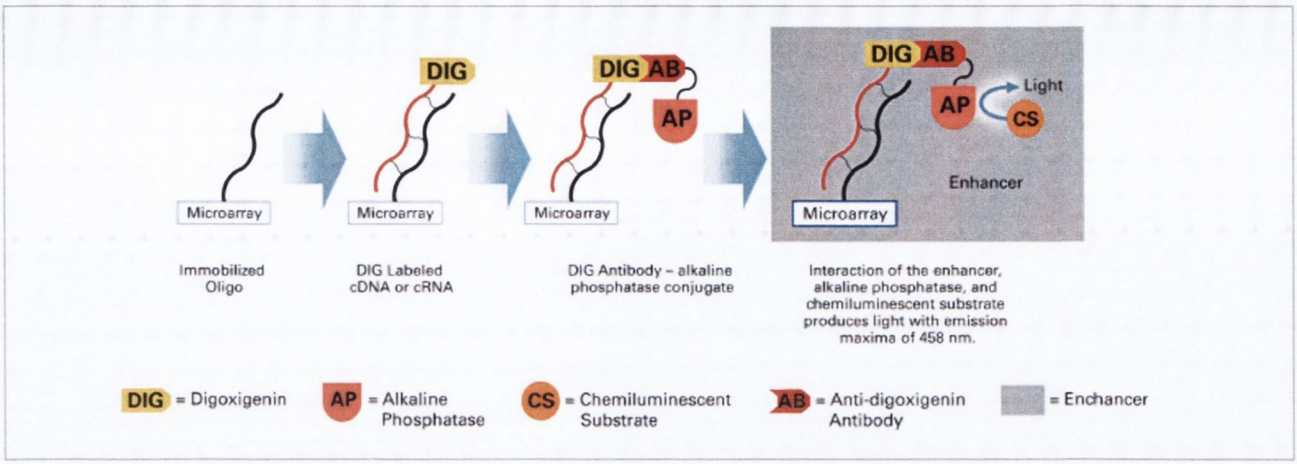



**Figure 4.1: Schematic of the labelling and chemiluminescence detection process used in the Expression Array System Microarray assay (Taken from www.appliedbiosystems.com).**

The 1700 Chemiluminescent Microarray Analyser (AB1700) images each microarray in fluorescent mode to grid, normalize and identify microarray features accurately even in the absence of bound gene expression products. The AB1700 then captures 2 "short" chemiluminescent images for coarse image analysis, quality assessment and quality control and 2 "long" chemiluminescent images for precise gene expression quantification.

### 4.4.4 Analysis of microarray data

After capturing array images and quantification of chemiluminescence for each probe, the AB1700 performs initial quality control analysis. Each array must pass this quality control to be included in subsequent analysis. For the purposes of this study, microarray data was then imported into a statistical package with the BioConductor R Programme called the Applied Biosystems Genome Survey Array Data Analysis program (ABarray), also referred to as ab1700gui. The ABarray performs normalisation and quality assessment on output from the AB1700. It can also be used to perform primary analysis and output lists of significant genes. Much of this analysis is carried out by a second program called R. This program applies filtering methods to identify genes that show differential expression across two study groups. Further comparisons were performed with Spotfire.

Names and functions of all genes of interest were downloaded from the PANTHER (**P**rotein **AN**alysis **TH**rough **E**volutionary **R**elationships) Classification System, available at www.pantherdb.org. Secondary analysis was also performed using PANTHER; the annotated gene list was analysed to identify molecular functions, biological processes and pathways that showed increased representation between the two study groups, either due to upregulation or downregulation of component genes.

In this study, there were six biological replicates for each of the three study groups. Of these six biological replicates, three were from three separate cell culture vials at passage 1 and three were from three separate cell culture vials at passage 2. Generating replicates is both time-consuming and expensive, but the more replicates, the better the statistics will ultimately be. Three replicates is the recommended minimum, however, in this study, there were 6 replicates.

#### 4.4.4.1 Primary analysis using ABarray

After normalisation and quality assessment, ABarray was used to conduct t tests and fold change analysis, looking for genes that showed significantly different levels of expression between two different cohorts (or classes). The classes examined were: a) murine embryonic stem cells, b) murine pluripotent teratocarcinoma cells and c) murine nullipotent teratocarcinoma cells. The output of this analysis was filtered to identify genes of interest.

##### *4.4.4.1.1 Class Comparison*

Class comparison is a method of data analysis that compares expression profiles from predefined specimen classes. The aims of class comparison are to determine whether the expression profiles differ among the classes and if so, to identify genes that are differentially expressed between classes. If possible, class comparison aims to develop a multivariate predictor of class membership based on level of expression of certain genes. Identification of differentially expressed genes was based on fold change analysis and significance testing.

*4.4.4.1.2 Significance*

The ABarray was used to perform a two-sample t test between all of the 2-class groups listed in table 4.2. This testing was preceded by a probe filtering procedure to remove undetected probes: probes in which the signal to noise ratio was less than 3. A threshold value, usually 50%, was selected for t test analysis; if a probe was not detected in more than 50% of the samples in a class, then it was listed as being undetectable in that class. For each probe, the t test looked at the mean and the variance of expression levels in class 1 compared with those in class 2 and calculated the probability of the observed difference in means occurring when the null hypothesis is true; i.e. when the mean of the two classes was equal. The t test in ab1700gui assumes unequal variance between classes. The output of this testing was a p value for each detected probe.

**Table 4.1: Groups used for analysis of microarray data**

| Analysis No. | Class 1 | Class 2 |
|---|---|---|
| 1 | Normal ES cells | Pluripotent teratocarcinoma |
| 2 | Normal ES cells | Nullipotent teratocarcinoma |

*4.4.4.1.3 Fold Change*

Fold change (FC) and $\log_2$(FC) for detected probes were also calculated. FC compares the expression level of the same gene in both classes and assesses how many fold up - or downregulated the gene is in one class compared with the other. In microarray studies, where number of replicates is low, but number of genes is high, there is a significant risk of experimental error. It is particularly difficult to estimate variance given a small number of replicates. To allow for this, it is recommended that only genes with a FC $\geq 2$ are included for analysis. This guards against low p values that arise from underestimation of variance (Knudsen, 2004).

*4.4.4.1.4 Gene Filtering*

Fold changes, p values from t tests and FDR values from multiple testing corrections were all used to filter gene lists from the 32,996 genes represented on each array to lists of between 100 to approximately 1000 genes that showed a significant difference in expression levels between 2 classes, e.g. in analysis 1, genes that were differentially expressed between mouse embryonic stem cells and pluripotent teratocarcinoma cells. Initially, basic filtering was automatically applied by ABarray, where flagged genes (that did not pass quality control) and those with a signal-to-noise ratio (S/N) of less than 3 were deemed undetectable and removed from further analysis. All control elements on the array such as hybridisation controls, RT-IVT controls, etc. were also removed, after being screened to ensure the data met the quality control requirements.

Preferentially, genes with an FDR $\leq 0.05$ were considered to show significant differential expression. However, if no or very few genes passed the Benjamini-Hochberg corrections, those with the smallest p value were evaluated, using $p \leq 0.01$. Finally, only genes with a FC $\geq$ 2 were included in subsequent analysis.

*4.4.4.1.5 Analysis using Spotfire*

Spotfire DecisionSite™ for Functional Genomics (Spotfire AB, Goteborg, Sweden) was used for more complex data comparisons. The two differentially expressed lists of genes obtained from the ABarray output were imported into Spotfire® prior to elimination of those genes with FDR $\leq 0.05$ and FC $\geq 2$ for further analysis. These two gene lists were compared using a venn diagram method to determine those genes common to both lists i.e. common to E v S undifferentiated and E v N undifferentiated. This produced a list of probe IDs and gene names only which were common to both gene lists. This list of probe IDs was then matched to fold change values from both original lists in R. Further refinement was performed by eliminating genes with a FC $\geq 2$ and FDR of $\leq 0.05$ (Reiner et al., 2003). The refined lists were then compared to determine only those genes significantly differentially expressed in both lists.

#### 4.4.4.2 Secondary analysis using PANTHER

The final list of genes obtained following ABarray and Spotfire analysis was imported into PANTHER as a list of probe IDs. PANTHER is a unique resource that classifies genes by their functions, using published scientific experimental evidence as well as sequence analysis to predict function even in the absence of direct experimental evidence. Gene products are classified into families and subfamilies of shared function, which are then categorized by molecular function and biological process. For an increasing number of proteins, interactions in biochemical pathways are also characterised. A key function of the PANTHER system is the ability to analyse gene lists for over- or under- represented biological processes.

#### 4.4.4.3 Cluster analysis

Cluster analysis is based on a metric for measuring distance between expression profiles. Clustering involves grouping samples or genes together according to some measure of similarity. Clustering can extract trends from raw data sets, but it does not give information about which genes are informative for discriminating among classes. There are two types of clustering: supervised and unsupervised clustering. Supervised clustering uses a set of example data to classify the rest of the data set. Unsupervised clustering, on the other hand, attempts to discover the natural groupings inside a data set without any input. Using the raw data from the AB microarray reader, cluster analysis was performed with the freely available clustering software, Cluster 3.0 and illustrated using the freely available Tree View 1.04, a cross-platform open source Java (programming language) based software application that allows interactive graphical analysis of the results from Cluster 3.0. This produced hierarchical clustering diagrams of the data in order to evaluate the degree of similarity between the expression profiles from all of the samples and to determine whether biological replicates and cell types grouped appropriately.

## 4.5 Results

### 4.5.1 RNA analysis

Quality and quantity of 18 total RNA samples was assessed by gel electrophoresis, spectrophotometry and using the Agilent 2100 Bioanalyser prior to use in expression array experiments (See Figure 4.2). Concentrations of the total extracted RNA and amount of RNA used in RT-IVT reactions are shown in Table 4.2.

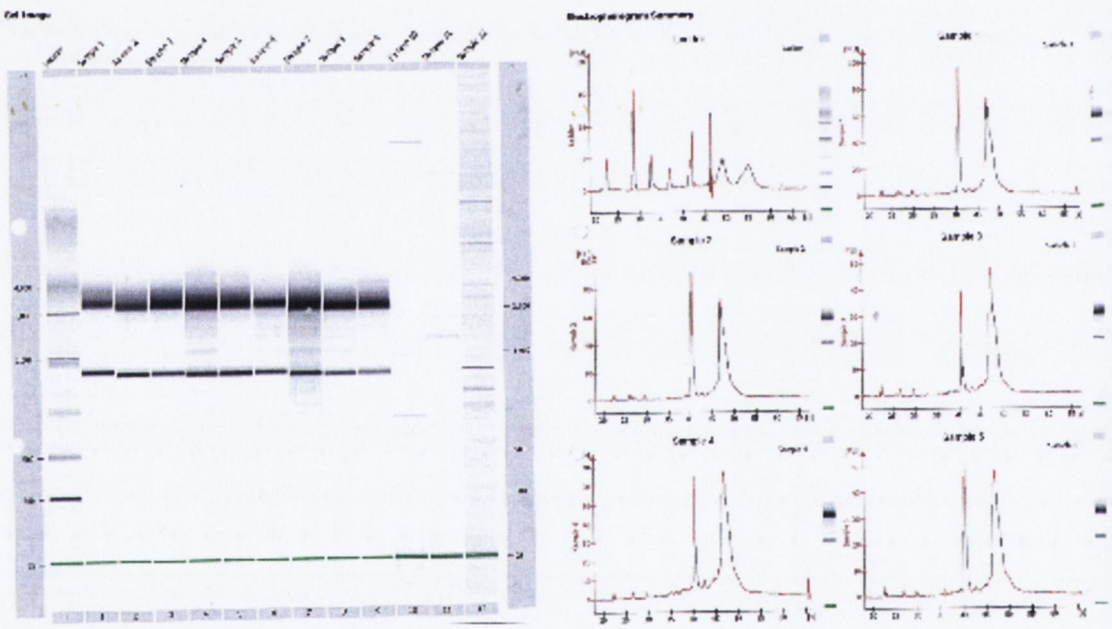


**Figure 4.2: Virtual gel and electropherograms for the S cell line timecourse.**

**Table 4.2: Concentrations (ng/ul) for samples used for Expression array analysis.**

| Sample | | Total RNA Conc (ug/ul) | Amount for RT-IVT (ug) | Sample | | Total RNA Conc (ug/ul) | Amount for RT-IVT (ug) |
|---|---|---|---|---|---|---|---|
| No | ID | | | No | ID | | |
| 1 | E1 | 0.710 | 2 | 10 | S4 | 2.906 | 2 |
| 2 | E2 | 0.870 | 2 | 11 | S5 | 1.856 | 2 |
| 3 | E3 | 0.492 | 2 | 12 | S6 | 0.490 | 2 |
| 4 | E4 | 1.424 | 2 | 13 | C1 | 1.092 | 2 |
| 5 | E5 | 1.996 | 2 | 14 | C2 | 1.152 | 2 |
| 6 | E6 | 1.619 | 2 | 15 | C3 | 0.568 | 2 |
| 7 | S1 | 0.568 | 2 | 16 | C4 | 1.184 | 2 |
| 8 | S2 | 1.644 | 2 | 17 | C5 | 0.760 | 2 |
| 9 | S3 | 1.138 | 2 | 18 | C6 | 1.588 | 2 |

### 4.5.2 cRNA analysis post RT-IVT

Post RT-IVT, DIG labelled cRNA was assessed qualitatively and quantitatively prior to array hybridisation. An example of a typical sample electrophoresed on an ethidium bromide-stained agarose gel is shown in Figure 4.3 with a comparison to four actual samples.

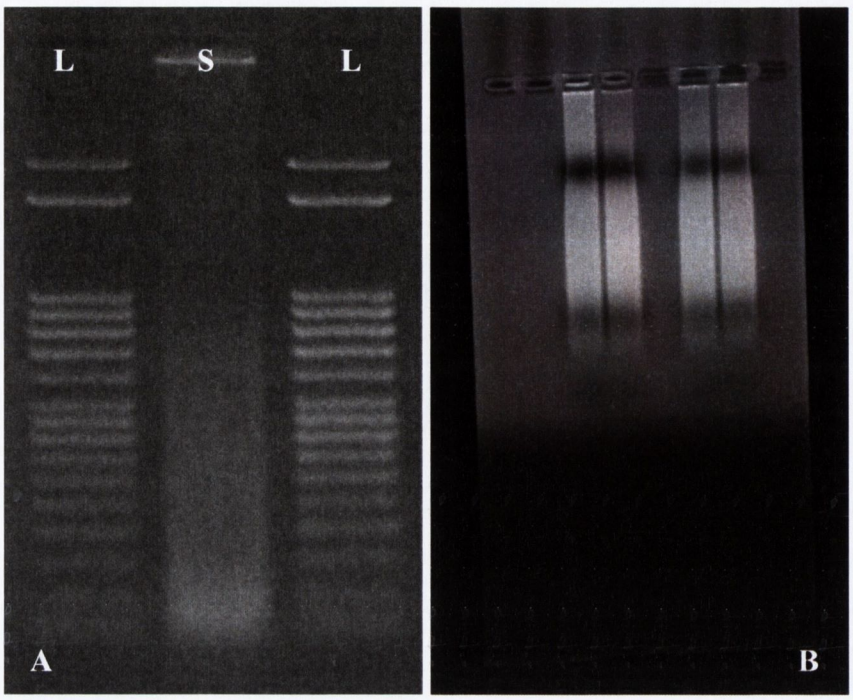



**Figure 4.3: A) Agarose gel electrophoresis of a cRNA sample - example. B) Agarose gel electrophoresis of 4 mouse cRNA samples. (L = size ladder, S = sample cRNA).**

cRNA concentration was also assessed prior to array hybridisation with 10μg of cRNA used per array (see Table 4.3 for details).

**Table 4.3: cRNA concentration of samples used for expression array analysis.**

| Sample | cRNA conc (µg/µl) | Amount for Array hybridisation (µg) | Sample | cRNA conc (µg/µl) | Amount for Array hybridisation (µg) |
|---|---|---|---|---|---|
| E1 | 0.1788 | 10 | S4 | 0.3828 | 10 |
| E2 | 0.1884 | 10 | S5 | 0.3492 | 10 |
| E3 | 0.4584 | 10 | S6 | 0.1572 | 10 |
| E4 | 0.1464 | 10 | C1 | 0.3300 | 10 |
| E5 | 0.1596 | 10 | C2 | 0.3012 | 10 |
| E6 | 0.3300 | 10 | C3 | 0.3756 | 10 |
| S1 | 0.1980 | 10 | C4 | 0.5700 | 10 |
| S2 | 0.4368 | 10 | C5 | 0.6240 | 10 |
| S3 | 0.2892 | 10 | C6 | 0.1824 | 10 |

### 4.5.3 Microarray image capture

Images of processed microarrays were captured and processed using the AB1700 system. Arrays were captured in two halves; an example is displayed in Figure 4.4.

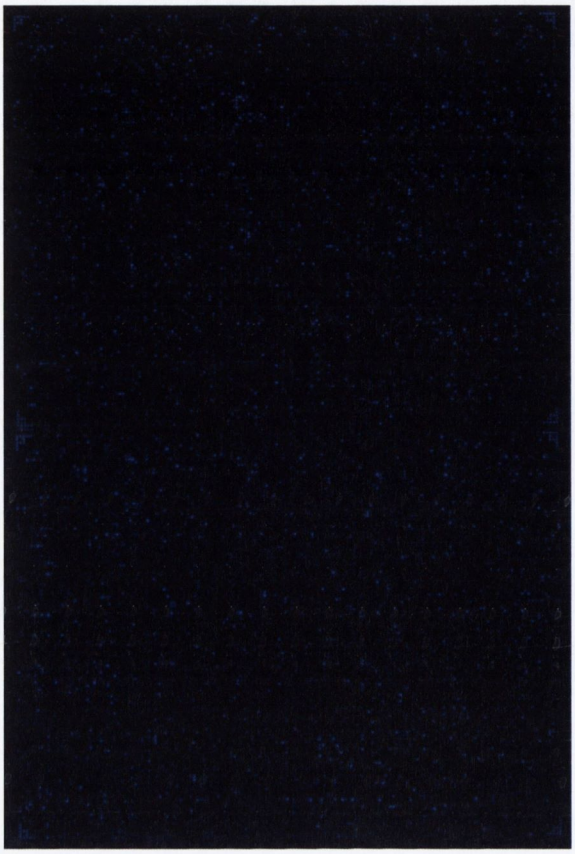


**Figure 4.4: Captured image of a microarray. The AB1700 captures the microarray image in two halves. These images have been merged for the purposes of this figure. (Take from www.appliedbiosystems.com).**

### 4.5.4 Statistical analysis

#### 4.5.4.1 Normalisation

A 5% trimmed mean was the main normalisation method used in this study; this method produces reliable results when analysing low signal intensity genes. Figure 4.4 shows MA plots after normalisation for the biological replicates E1 - E6, S1 – S6, C1 – C6. MA plots for replicates are symmetrical around zero and show no appreciable difference. The three replicates from each passage show closer similarities than a comparison of a replicate from each of the two passages, for example comparison of E1 and E6 shows a wider variation around 0 but the correlation is still high at 96.4%.

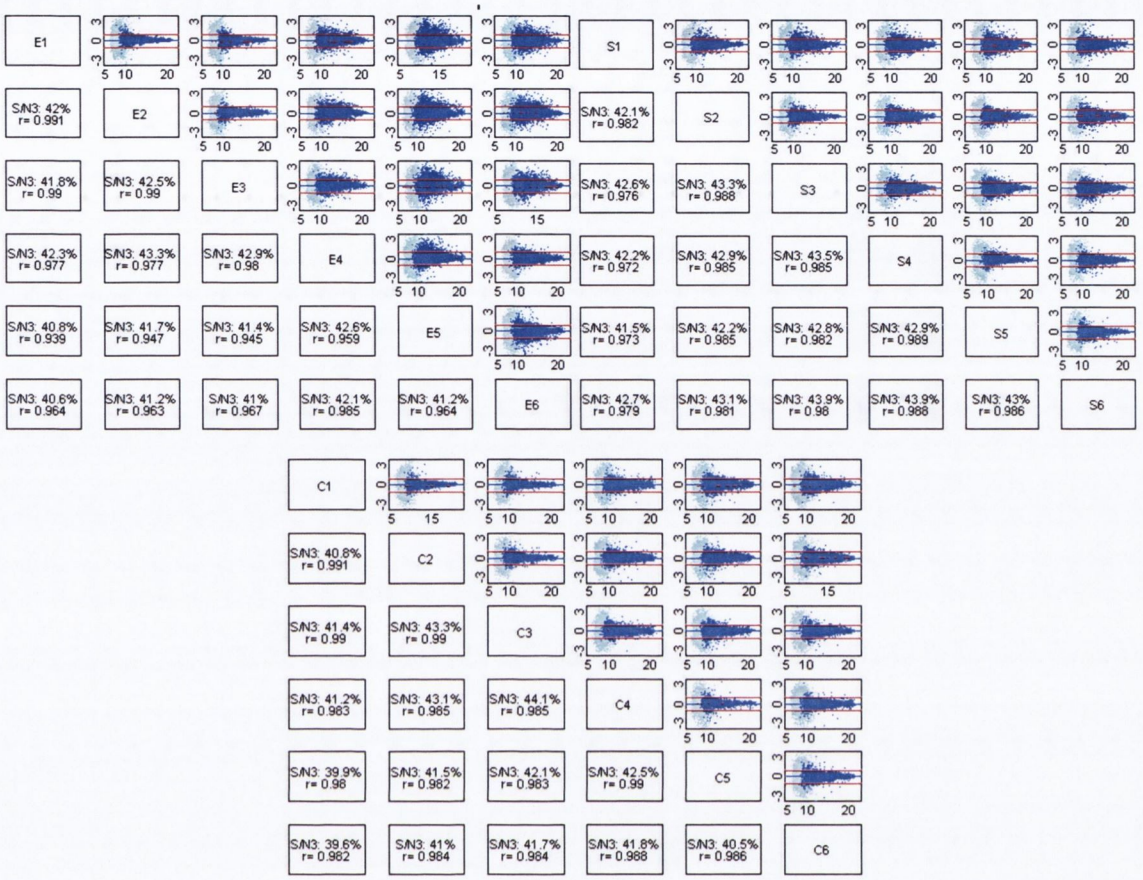



**Figure 4.5: MA plots after normalisation for the biological replicates E1 - E6, S1 – S6, C1 – C6.**

### 4.5.4.2 Filtering genes

Step-wise filtering of gene probes was performed for all analyses as follows:

1) During normalisation in ABarray, probes with a signal-to-noise ratio of less than 3 (S/N<3) in all arrays were eliminated.

2) Detectable genes were then analysed using a t test in ab1700gui and a correction for multiple testing. Genes with an FDR ≤ 0.05 were considered to show significant differential expression between two classes of analysis. Genes that were not significantly

differentially expressed after t test or adjustment for multiple comparisons were eliminated.

3) Genes less than two-fold up- or down-regulated (FC < 2) between the two classes were eliminated.

#### 4.5.4.3 Analysis of normal ES cells v malignant teratocarcinoma cells

**Analysis 1:** E (mES samples) v S (pluripotent teratocarcinoma samples)

A total of 12 samples were used for this analysis (6 x normal ES samples and 6 x pluripotent teratocarcinoma samples) assigned as either benign or malignant. After normalisation and quality control, 16,994 probes were considered for further analysis.

**Analysis 2:** E (mES samples) v C (nullipotent teratocarcinoma samples)

A total of 12 samples were used for this analysis (6 x normal ES samples and 6 x nullipotent teratocarcinoma samples) assigned as either benign or malignant. After normalisation and quality control, 16,976 probes were considered for further analysis.

**Analysis 3:** Comparison of both lists in Spotfire:

The probe IDs from the two lists above were imported into Spotfire. Using a venn diagram method, these two lists of 16,994 and 16,976 probes were compared.

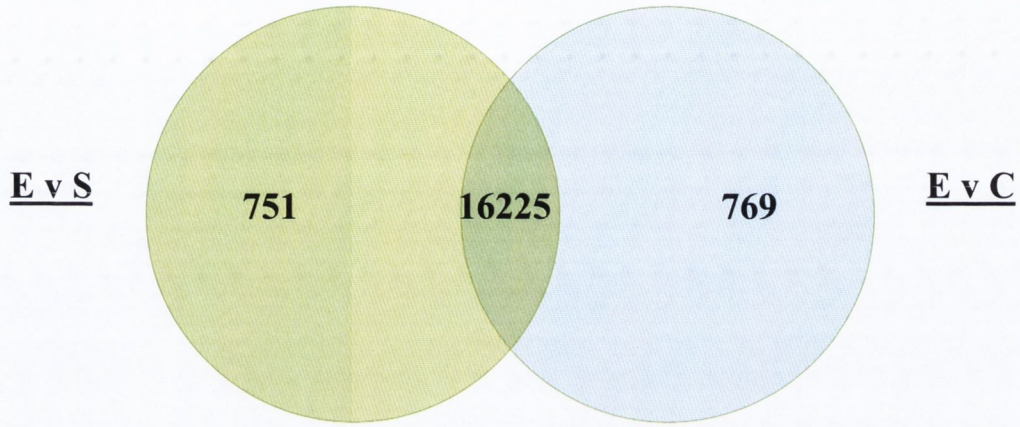



**Figure 4.6: Venn diagram showing number of genes common to both lists (16,225) and exclusive to each list (751 and 769). E = mES cell line, S = SCC-PSA1 cell line, C = Nulli-SCC1 cell line.**

16,225 genes were common to both lists with 751 genes unique to E v S and 769 genes unique to the E v C comparison. In order to examine all the malignant events, the genes common to both lists (16,225) were focused on and matched to the original E v S and E v C fold change values to perform further analysis. Further gene filtering was performed at this stage with only genes with FC $\geq 2$ and FDR of $\leq 0.05$ considered further. This produced a list of 1,170 genes, differentially up and downregulated between ES cells (E) and their malignant counterparts (S and C) as outlined in Figure 4.6.

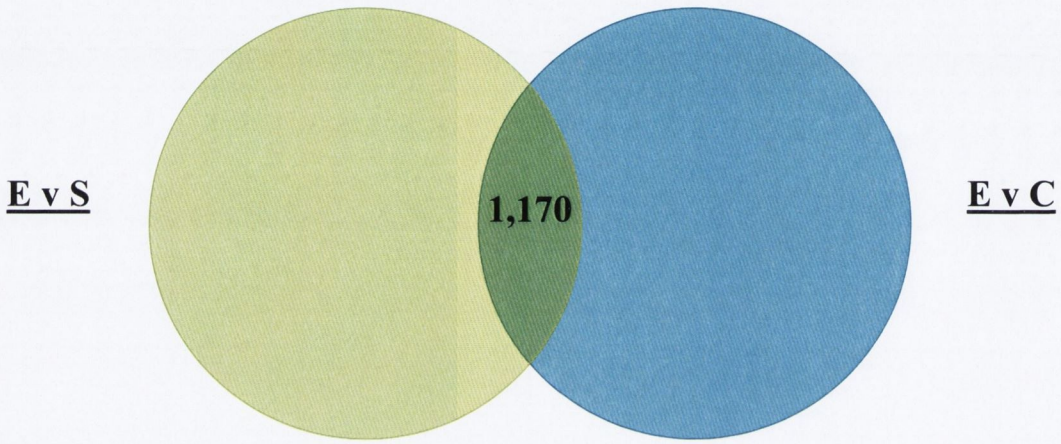



**Figure 4.7: Venn diagram showing number of genes common to both gene lists with FDR < 0.05 and fold change > 2.**

Of these genes, 50-55% was downregulated in malignant cells and 45-49% was upregulated in malignant cells. 43.2% of this list of genes was unknown or unnamed. The probe IDs of the 1,170 common differentially expressed genes were imported into PANTHER to look at their functional properties. Gene IDs were found for 1,108 of the 1170 probe set.

The challenge, as with many microarray studies, was how to get meaningful information from such a large number of significant genes.

To obtain the maximal amount of useful information:

a) The most highly upregulated genes in malignant cell lines compared with benign when the 1,170 gene list was matched up to both original lists were obtained. Fold change values are shown for both lists for each gene (Table 4.4).

b) The most highly downregulated genes in malignant cell lines compared with benign when the 1,170 gene list was matched to both original lists were obtained (Table 4.5).

c) The 1,170 gene list was imported into PANTHER to obtain information on the pathways, biological processes and molecular functions significantly over-represented in this list listed in Table 4.6.

**Table 4.4: Top 10 most highly upregulated genes in murine nulli (C) and pluripotent (S) malignant teratocarcinoma cells compared with mES (E) cells.**

| | Genes upregulated in Malignant cell lines<br>C v E comparison<br>Gene symbol | C v E (FC) | S v E (FC) |
|---|---|---|---|
| *1.* | ***Serpina3m, serine (or cysteine) proteinase inhibitor, clade A, member 3M*** | 90.9 | 7.3 |
| *2.* | ***Pramel6, preferentially expressed antigen in melanoma like 6*** | 76.9 | 0.2 |
| *3.* | ***H2-Eb1, histocompatibility 2, class II antigen E beta*** | 58.8 | 2.7 |
| *4.* | ***Xlr5, X-linked lymphocyte-regulated 5*** | 52.6 | 4.7 |
| *5.* | ***Tex13, testis expressed gene 13*** | 47.6 | 14.5 |
| *6.* | ***Adm2, adrenomedullin 2*** | 37.0 | 17.5 |
| *7.* | ***Serpina3k, serine (or cysteine) proteinase inhibitor, clade A, member 3K*** | 35.7 | 3.6 |
| *8* | ***Bhmt, betaine-homocysteine methyltransferase*** | 31.3 | 2.5 |
| *9* | ***Ccna1, cyclin A1*** | 27.8 | 10.0 |
| *10.* | ***Hoxd8, homeo box D8*** | 25.0 | 5.5 |
| | **S v E comparison**<br>**Gene symbol** | **C v E (FC)** | **S v E (FC)** |
| *1.* | ***Hkdc1, hexokinase domain containing 1*** | 0.2 | 58.8 |
| *2.* | ***Hs3st1, heparan sulfate (glucosamine) 3-O-sulfotransferase 1*** | 0.2 | 55.6 |
| *3.* | ***Klk6, kallikrein 6*** | 11.5 | 47.6 |
| *4.* | ***Aqp8, aquaporin 8*** | 0.3 | 41.7 |
| *5.* | ***Glipr1, GLI pathogenesis-related 1 (glioma)*** | 0.3 | 40.0 |
| *6.* | ***Prg1, proteoglycan 1, secretory granule*** | 0.2 | 40.0 |
| *7.* | ***Galgt2, UDP-N-acetyl-alpha-D-galactosamine:(N-acetylneuraminyl)-galactosyl-N-acetylglucosaminylpolypeptide-beta-1, 4-N-acetylgalactosaminyltransferase*** | 4.0 | 38.5 |
| *8* | ***F2, coagulation factor II*** | 2.9 | 33.3 |
| *9* | ***Pthr1, parathyroid hormone receptor 1*** | 0.1 | 33.3 |
| *10.* | ***Ctsh, cathepsin H*** | 0.1 | 27.8 |

**Table 4.5: Top 10 most highly downregulated genes in murine nulli (C) and pluripotent (S) malignant teratocarcinoma cells compared with mES (E) cells.**

| | Genes downregulated in Malignant cell lines<br>C v E comparison<br>Gene symbol | C v E (FC) | S v E (FC) |
|---|---|---|---|
| *1.* | ***Dcn, Decorin*** | 0.0006 | 0.4859 |
| *2.* | ***Tnc, Tenascin C*** | 0.0014 | 0.2811 |
| *3.* | ***Tm4sf1, transmembrane 4 superfamily member 1*** | 0.0016 | 0.3775 |
| *4.* | ***Ccl2, chemokine (C-C motif) ligand 2*** | 0.0017 | 0.2986 |
| *5.* | ***Eif2s3y, eukaryotic translation initiation factor 2, subunit 3, structural gene Y-linked*** | 0.0022 | 0.0026 |
| *6.* | ***Col3a1, procollagen, type III, alpha 1*** | 0.0025 | 0.2905 |
| *7.* | ***S100a4, S100 calcium binding protein A4*** | 0.0025 | 0.2820 |
| *8* | ***Nrp1, neuropilin 1*** | 0.0036 | 0.2036 |
| *9* | ***Ptn, pleiotrophin*** | 0.0066 | 0.1647 |
| *10.* | ***Col6a3, procollagen, type VI, alpha 3*** | 0.0097 | 0.4717 |
| | **S v E comparison**<br>**Gene symbol** | **C v E (FC)** | **S v E (FC)** |
| *1.* | ***Eif2s3y, eukaryotic translation initiation factor 2, subunit 3, structural gene Y-linked*** | 0.002 | 0.003 |
| *2.* | ***Ube1y1, ubiquitin-activating enzyme E1, Chr Y 1*** | 0.023 | 0.017 |
| *3.* | ***Tcstv1, 2-cell-stage, variable group, member 1*** | 0.257 | 0.022 |
| *4.* | ***Tcstv3, 2-cell-stage, variable group, member 3*** | 0.132 | 0.030 |
| *5.* | ***Ctla2a, cytotoxic T lymphocyte-associated protein 2 alpha*** | 0.022 | 0.031 |
| *6.* | ***Hoxb1, Homeobox b1*** | 0.079 | 0.058 |
| *7.* | ***Uty, ubiquitously transcribed tetratricopeptide repeat gene, Y chromosome*** | 0.063 | 0.063 |
| *8* | ***Mafb, v-maf musculoaponeurotic fibrosarcoma oncogene family, protein B (avian)*** | 0.066 | 0.065 |
| *9* | ***Hoxb2, Homeobox b2*** | 0.123 | 0.068 |
| *10.* | ***Olig3, oligodendrocyte transcription factor 3*** | 0.086 | 0.071 |

Pathway analysis:

Within PANTHER, the most significantly over represented pathways, molecular functions and biological processes in this list of 1,170 genes were identified by comparison to a reference list of mouse AB1700 genes (Table 4.6).

**Table 4.6: Top 5 pathways, molecular functions and biological processes over-represented in the 1170 gene list. Note: Number in brackets indicates number of genes within this group.**

| No. | Pathways | P value | Molecular Functions | P value | Biological Processes | P value |
|---|---|---|---|---|---|---|
| 1 | Oxidative stress response (9) | 0.83 | Transcription factors (114) | 0.0006 | Developmental processes (146) | $3.69x10^{-7}$ |
| 2 | Apoptosis signalling pathway (8) | 1 | Reverse transcriptase (6) | 0.052 | mRNA transcription (119) | $7.8x10^{-6}$ |
| 3 | Angiogenesis (15) | 1 | Extracellular matrix structural proteins (11) | 0.128 | mRNA transcription regulation (94) | 0.0001 |
| 4 | Wnt signalling pathway (25) | 1 | Extracellular matrix (27) | 0.144 | Neurogenesis (51) | 0.0007 |
| 5 | TGF-Beta signalling pathway (14) | 1 | Homeobox transcription factors (21) | 0.277 | Ectoderm development (55) | 0.0008 |

Interestingly, none of the pathways found is statistically significantly over represented in this gene group despite a number of these pathways having been reported in the literature to be significant in stem cell biology, for example Wnt and TGF-β signalling pathways. Of the molecular functions, transcription factor was the most significant and contains 114 genes. The biological processes group appears to have the most significant information with the top 5 over represented processes all statistically significant. The processes involved, i.e., developmental processes, mRNA transcription and transcriptional regulation, neurogenesis and ectoderm development are not unexpected biological processes in the area of stem cells which was a reassuring result.

#### 4.5.4.4 Cluster analysis

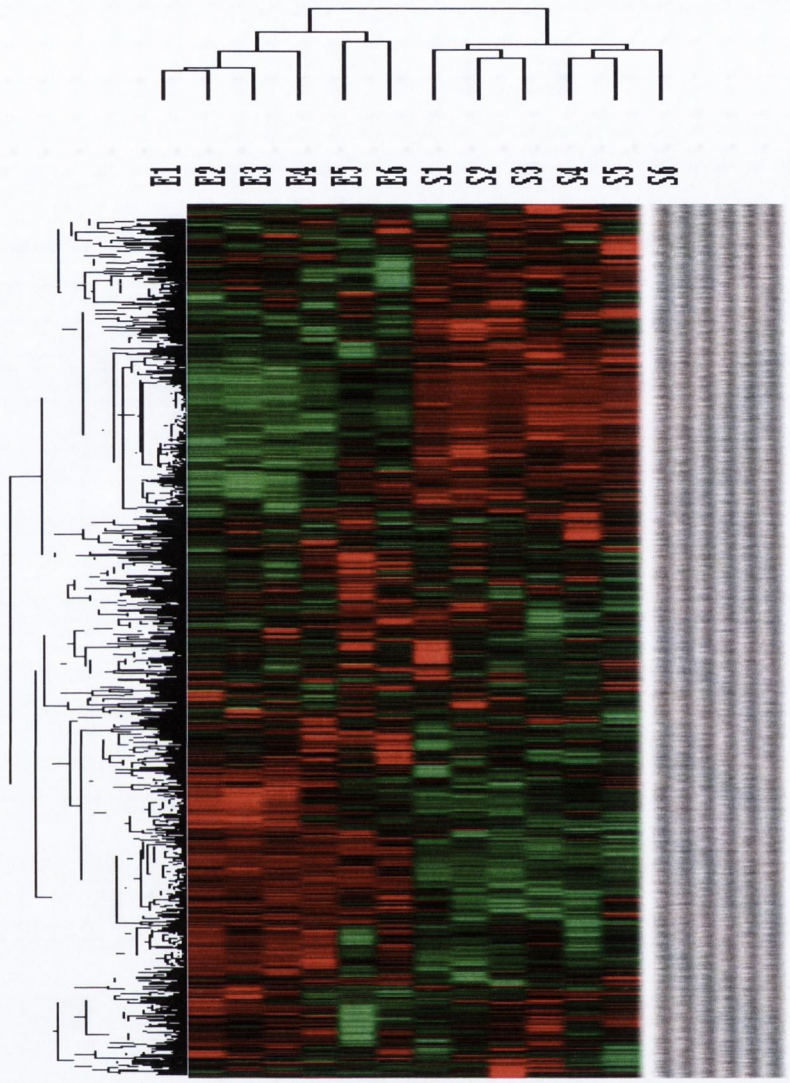



**Figure 4.8: Unsupervised hierarchical clustering of E v S cell lines using Gene Cluster 3.0.**

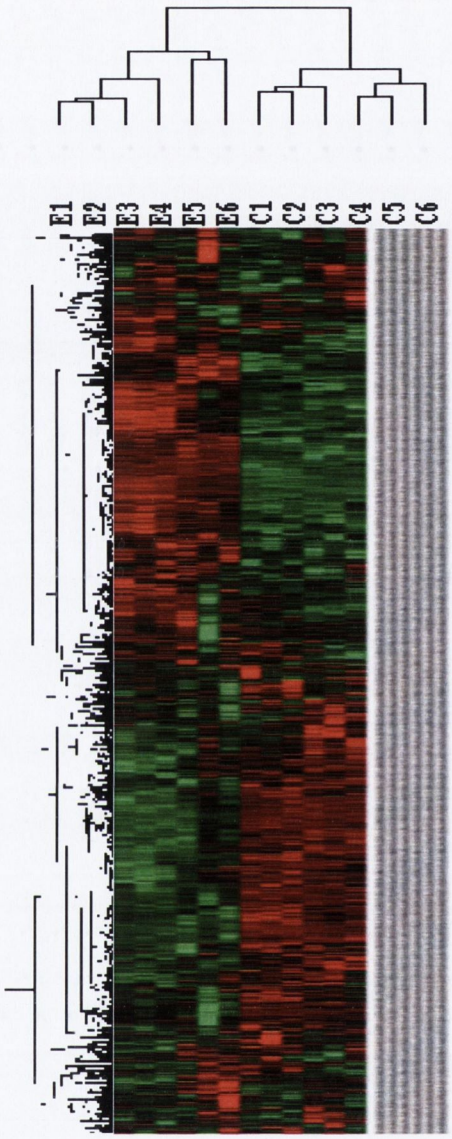



**Figure 4.9: Unsupervised hierarchical clustering of E v C cell lines using Gene Cluster 3.0.**

Hierarchical clustering was performed to validate the reproducibility and overall variation of the data. Unsupervised hierarchical clustering using Gene Cluster 3 and Tree View 1.04 to illustrate the results shows the E cell line clustering together and the respective S and C cell lines clustering together as expected. The results obtained indicate that the overall inter experimental variation was low with all six E cell line replicates clustering together and separately from both of the malignant cell lines (S and C) as expected. Interestingly, the three replicates from each passage tend to cluster together demonstrating some small differences

between samples from passage 1 (E/S/C1 – 3) and passage 2 (E/S/C4-6). This is supported by the MA plots where the correlation was tighter between the replicates of each passage.

### 4.5.5 Validation of Microarray results

Validation of microarray results was performed using real-time PCR. The reasons for undertaking validation experiments are:

1. To verify that the observed changes are reproducible in a larger number of samples.
2. To verify that array results are not the result of problems inherent to the array technology.

Microarrays are known to be an excellent tool for initial target discovery but there are many sources of variability that may affect results including variability from laboratory to laboratory, user to user and platform to platform. Therefore, it is essential to use independent means to verify that the genes of interest are truly differentially expressed, and to what extent.

For this study, 47 targets were validated and are listed in detail in Chapter 2, Table 2.11.

**Table 4.7: 47 targets validated in the undifferentiated benign v malignant comparisons with their values from the microarray and TaqMan® experiments. (Yellow = downregulated, grey = upregulated, blue = different in each group, no fill = not significant, Not found = not differentially expressed, TQ = TaqMan®, FC = Fold change, FDR = false discovery rate).**

| | | E v C | | | E v S | | |
|---|---|---|---|---|---|---|---|
| | | TQ FC | Array FC | FDR | TQ FC | Array FC | FDR |
| 1 | Adamts5 | 90.296 | Not found | | 0.77 | 0.39 | 0.1 |
| 2 | AFP | | Not found | | 0.033 | 0.306 | 0.06 |
| 3 | Ccnd1 | 24.302 | 8.565 | 0.002 | 5.592 | 2.446 | 0.013 |
| 4 | Cdkn1c | 67.932 | 31.947 | 5.31E-05 | 25.704 | 11.625 | 0.0004 |
| 5 | Cited1 | 0.442 | 0.185 | 0.0005 | 0.112 | 0.107 | 0.0001 |
| 6 | Dcn | 57021 | 1631.056 | 7.04E-07 | 6.616 | 2.058 | 0.025 |
| 7 | Dscam | 0.658 | Not found | | 0.208 | 0.562 | 0.13 |
| 8 | Egr4 | 38.442 | 6.611 | 0.002 | 11.583 | 4.38 | 0.017 |
| 9 | Eif2s3y | Undet in N | 460.098 | 4.12E-05 | Undet in S | 387.854 | 6.88E-07 |
| 10 | Fgf5 | 36.405 | 5.966 | 0.007 | 3.124 | 3.017 | 0.008 |
| 11 | Foxj1 | 0.989 | 0.394 | 0.0001 | 0.526 | 0.46 | 0.0005 |
| 12 | Fzd1 | 29.519 | 9.926 | 0.0006 | 12.071 | 3.858 | 0.006 |
| 13 | Galgt2 | 0.093 | 0.251 | 0.0004 | 0.005 | 0.026 | 2.05E-06 |
| 14 | Gata1 | 0.071 | 0.064 | 0.0002 | 0.417 | Not found | |
| 15 | Gata6 | 11.757 | 1.972 | 0.063 | 0.058 | 0.066 | 1.63E-06 |
| 16 | Gdf15 | 0.07 | 0.054 | 1.87E-07 | 0.381 | 0.29 | 1.18E-06 |
| 17 | Hhip | 0.422 | 0.098 | 0.0003 | 2.785 | 0.751 | 0.538 |
| 18 | Hoxb1 | 155.885 | 12.721 | 2.63E-05 | 490.382 | 17.134 | 1.42E-05 |
| 19 | Hoxd1 | 0.383 | 0.495 | 0.038 | 0.744 | Not found | |
| 20 | Hoxd9 | 0.027 | 0.06 | 3.43E-05 | 0.182 | Not found | |
| 21 | Igfbp5 | 4.467 | Not found | | 0.806 | 0.851 | 0.595 |
| 22 | Klf2 | 0.61 | 0.277 | 4.00E-03 | 2.26 | 1.446 | 0.233 |
| 23 | Klk6 | 0.121 | 0.087 | 0.003 | 0.034 | 0.021 | 0.0006 |
| 24 | Lefty1 | 36.985 | 9.286 | 3.00E-04 | 20.438 | 11.903 | 8.83E-05 |
| 25 | Lifr | 0.638 | 0.338 | 0.006 | 0.303 | 0.117 | 0.003 |
| 26 | Mafb | 18.676 | 15.184 | 7.00E-04 | 15.332 | 15.407 | 0.0007 |
| 27 | Mmp2 | 54.238 | 12.157 | 6.77E-05 | 2.881 | 1.138 | 0.645 |
| 28 | Myc | 47.845 | 12.295 | 2.000E-04 | 16.416 | 2.698 | 0.027 |
| 29 | Nanog | 0.494 | 0.332 | 5.000E-04 | 1.914 | 1.056 | 0.756 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 30 | **Notch4** | 0.263 | 0.597 | 3.000E-03 | 0.472 | 0.054 | 2.44E-06 |
| 31 | **Oct4** | 1.133 | 0.697 | 5.000E-03 | 1.545 | *1.052* | *0.539* |
| 32 | **Olfr1450** | Undet in N | Not found | | Undet in S | Not found | |
| 33 | **Olig3** | 30.941 | 11.665 | 6.000E-04 | 123.425 | 14.02 | 0.001 |
| 34 | **Pramel6** | 0.02 | 0.013 | 4.57E-07 | 48.532 | 6.307 | 0.0009 |
| 35 | **Rab15** | 0.713 | 0.362 | 6.000E-04 | 0.551 | 0.413 | 0.002 |
| 36 | **Rhoj** | 354.519 | 29.527 | 5.79E-05 | 11.03 | 6.254 | 0.001 |
| 37 | **Scml2** | 0.139 | 0.087 | 0.0003 | 1.34 | Not found | |
| 38 | **Slamf9** | 0.206 | 0.048 | 5.29E-05 | 5.954 | Not found | |
| 39 | **Sox17** | 6.215 | 4.431 | 2.00E-03 | 0.045 | 1.453 | 4.67E-06 |
| 40 | **Sox2** | 1.393 | 0.656 | 7.00E-03 | 1.495 | *1.002* | *0.99* |
| 41 | **Tcstv1** | Undet in N | 7.626 | 3.00E-04 | 5.752 | 46.081 | 2.96E-07 |
| 42 | **Tert** | 0.298 | 0.242 | 3.00E-04 | 0.665 | 0.415 | 0.009 |
| 43 | **Tex12** | 0.355 | Not found | | 0.126 | 0.233 | 7.54E-06 |
| 44 | **Timp2** | 393.914 | 194.546 | 2.65E-06 | 3.255 | 1.238 | 0.41 |
| 45 | **Tnc** | | 732.976 | 4.56E-05 | | 3.557 | 0.027 |
| 46 | **Tnfrsf9** | | Not found | | | 0.881 | 0.633 |
| 47 | **Wnt6** | 1.392 | 0.446 | 1.00E-03 | 0.798 | 0.431 | 0.001 |

Looking specifically at the three most well known and discussed stem cell markers, Nanog, Oct4 and Sox2, none is significantly differentially expressed between E and S cell lines, a not unexpected result as both cell lines are in their undifferentiated state and awaiting stimulation to differentiate. In the comparison of E and C cell lines, there are some changes with a significant FDR but fold changes are $< 2$ and $> 0.5$ for Oct4 and Sox2 with consistent results using both microarray and TaqMan® analysis. Nanog is slightly different and with a fold change $< 0.5$ and FDR $< 0.05$ is significantly downregulated using both microarray and TaqMan® results.

**Table 4.8: Comparison of common stem cell markers in 'benign' v 'malignant' stem cell.**

| | E v C | | | E v S | | |
|---|---|---|---|---|---|---|
| | TQ FC | Array FC | FDR | TQ FC | Array FC | FDR |
| Nanog | 0.494 | 0.332 | 5.000E-04 | 1.914 | 1.056 | 0.756 |
| Oct4 | 1.133 | 0.697 | 5.000E-03 | 1.545 | 1.052 | 0.539 |
| Sox2 | 1.393 | 0.656 | 7.00E-03 | 1.495 | 1.002 | 0.990 |

As Nanog is a gene required for the maintenance of self renewal, a significant difference in expression would not be expected. While the levels of Nanog have not been shown to be critical to the self-renewal state, elevated levels of Nanog would confer constitutive self renewal abilities in a given cell line (Darr et al., 2006). Since the C cell line is a nullipotent cell line and maintains it in a continuous state of self renewal with no differentiation possible, this may explain up regulation and increased activity in the C cell line and the converse downregulation in the E cell line.

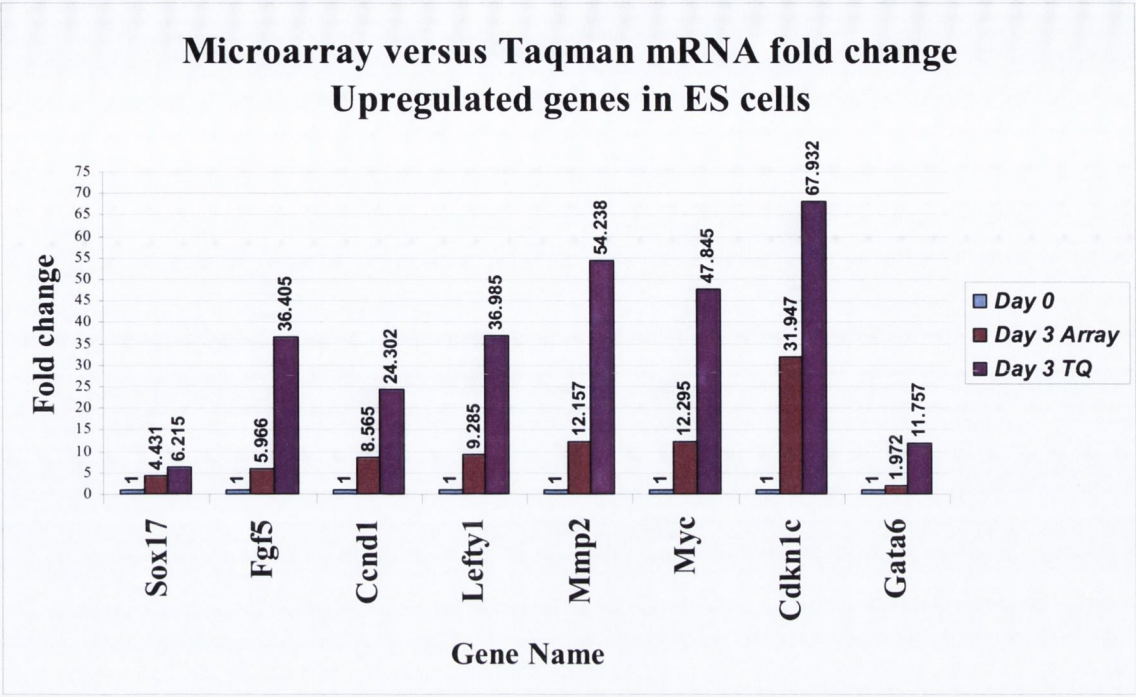



**Figure 4.10: Selected upregulated genes in ES cells compared with C cells with a comparison of microarray and TaqMan® mRNA fold change values. (TQ = TaqMan®).**

A comparison of selected upregulated genes in the E cell line between the microarray and TaqMan® fold change values shows consistency with TaqMan® analysis giving a higher fold change value compared with the microarray in each case. This may be due to TaqMan® methods being more accurate.

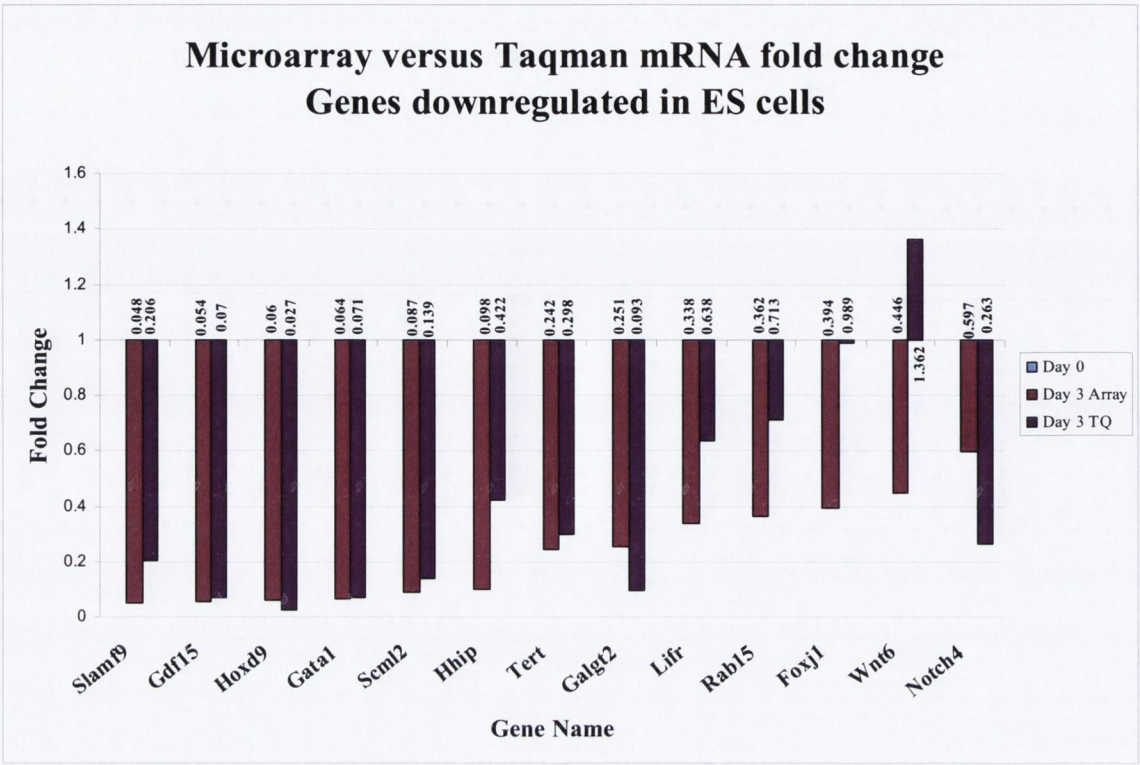
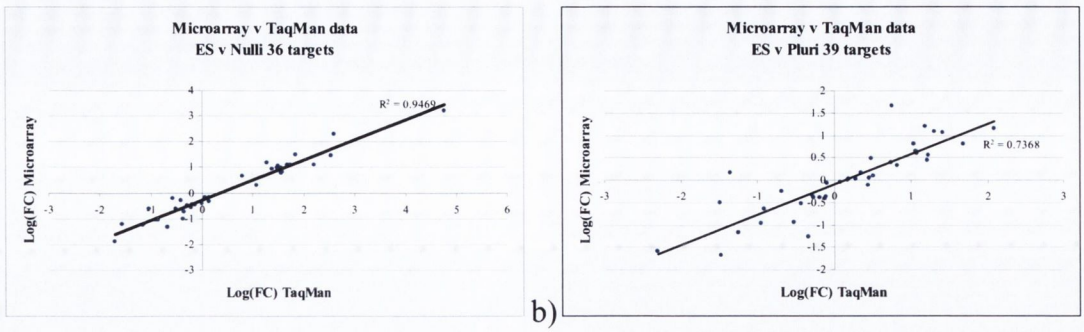



**Figure 4.11: Selected genes downregulated in ES cells compared with C cells with their fold change values from microarray and TaqMan® analysis.**

Looking at the downregulated genes, microarray and TaqMan® values are more similar with Wnt6, one of the Wnt genes involved in primitive/parietal endoderm differentiation (Krawetz et al., 2007), showing a disparate result between the microarray and TaqMan® result. Referring to Table 4.7, microarray analysis demonstrated down regulation in E cell lines compared with both malignant cell lines with a FDR $< 0.05$ in both cases. This down regulation was not confirmed by TaqMan® analysis. This may be due to a difference between the probe used in the TaqMan® assay and that used in the microarray.




a) b)

**Figure 4.12: Validation of microarray data. a) Log (FC) TaqMan® *vs* Log (FC) Microarray in ES *vs* Nulli comparison. Correlation coefficient = 0.97, slope = 0.78. b) Log (FC) TaqMan® *vs* Log (FC) Microarray in ES *vs* Pluri comparison. Correlation coefficient = 0.86, slope = 0.68.**

Looking at an overall comparison of microarray versus TaqMan® analysis, shows a good correlation of results with the Pearson correlation coefficient ranging from 0.86 to 0.97.

## 4.6 Discussion

Using cRNA microarrays with independent validation with real-time RT-PCR, we performed a unique comparison of normal mES cells with their malignant counterpart, murine teratocarcinomas. To incorporate as many malignant events as possible, this study used two teratocarcinoma cell lines, one being pluripotent, thus capable of differentiation, and one being nullipotent, thus incapable of differentiation and therefore a more aggressive tumour. The direct comparison of the cancer stem cell of teratocarcinoma to its normal counterpart should provide a vast array of potential biomarkers and highlight differences between the two that may provide a therapeutic window. ES cells have been well characterised in both mouse and humans with previous studies comparing them to adult stem cells, including haematopoietic and neural stem cells (Ivanova et al., 2002; Ramalho-Santos et al., 2002), or focusing on the comparison of the undifferentiated to differentiated state in order to obtain information on stem cells thus leaving a scarcity of information on the direct comparison of a stem cell and its malignant counterpart. In this study, the fundamental differences between a normal mouse embryonic stem cell and its malignant counterpart in their resting states were examined in an attempt to uncover the basic underlying differences between a normal stem cell and its malignant equivalent. The challenge was to pinpoint the differences between cancer stem cells

and normal stem cells as identification of these distinctive differences will provide novel targets for future cancer therapies (Bjerkvig et al., 2005). Indeed, the actual targeting of cancer stem cells as a therapeutic option has now been put forward by a number of groups (Reya et al., 2001; Sell, 2004; Huff et al., 2006) and the ultimate success of a treatment may be dependent on the elimination of the cancer stem cell in any given tumour (Zhang et al., 2006). However, a difficulty that may arise with the elimination of cancer stem cells is the eradication of their normal counterparts (Huff et al., 2006). This study's comparison of a normal stem cell and its malignant counterpart aims to overcome this predicament and thus provide more effective methods of detection and treatment of both teratocarcinoma and other tumours.

The most significant pathway represented in the data was that of oxidative stress, a mechanism that protects the cell from ageing and thus a property required by stem cells to retain their enduring status within the body (Koestenbauer et al., 2006). Ramalho-Santos *et al.* found a similar theme of cells under stress in their comparisons of embryonic and adult stem cells (Ramalho-Santos et al., 2002). Among the genes involved in the oxidative stress response were the dual specificity phosphatases (DUSPs), a subclass of protein tyrosine phosphatases uniquely able to hydrolyse the phosphate ester bond on both a tyrosine and a threonine or serine residue on the same protein (Theodosiou et al., 2002). The expression of some of the DUSPs differs in cancer stem cells in our study compared with the reported expression of these molecules in other malignancies (Hoornaert et al., 2003; Yu et al., 2007). Indeed, DUSP9 has not been reported previously in malignancy. DUSPs have also been investigated for their use as antineoplastic agents as illustrated by Lyon et al. in various cancer and Alzheimer's disease (Lyon et al., 2002). TGF-β signalling pathways were also found by Ramalho-Santos et al. to have an important role in stem cells and again this pathway was over-represented in our data (Ramalho-Santos et al., 2002). TGF-β signalling is involved in a wide range of cell fate decisions and cellular processes (Stewart et al., 2006). Apoptosis and angiogenesis were also significantly over-represented in the data presented in this chapter and correlated with two of the documented hallmarks of cancer (Hanahan et al., 2000).

Genes involved in the extracellular matrix were found to be over-represented in this study's dataset. Indeed, Decorin is one of the most down-regulated genes in our malignant population compared with the normal mES population. Decorin is a member of a family of small leucine-rich proteoglycans and is involved in a number of cellular processes including matrix

assembly, fibrillogenesis and the control of cell proliferation (Iozzo, 1998). Its expression has also been shown to be differentially down-regulated in hepatocellular, lung and ovarian tumours (Seidler et al., 2006). Thus, this finding of reduced Decorin expression in malignancy is reaffirmed in this study.

## 4.7 Conclusion

This chapter has produced a unique transcriptome profile of the cancer stem cell using a mouse model of ES and both pluri- and nullipotent EC cells which has not previously been performed. The profile consists of a list of 1,170 differentially expressed genes from which functional analysis and highly significant genes were obtained. As well as outlining some of the typical pathways, biological processes and molecular functions associated with malignancy and included in the hallmarks of cancer thus confirming the findings of other authors in the current literature on stem cells, many novel differences have been highlighted which may be of potential significance in the search for novel therapeutic strategies against cancer and the specific targeting of the cancer stem cell. The oxidative stress response is shown to be significant in the cancer stem cell state with many novel oxidative stress genes such as the Dusps found to be significantly differentially expressed. In addition, the importance of signalling molecules and the extracellular matrix or stem cell 'niche' was highlighted. Decorin and Tenascin C, two extracellular matrix genes, were two of the most significantly differentially regulated in this comparison and thus considered worthy of future study along with further investigation of the role of the extracellular matrix in cancer stem cell self-renewal and proliferation.

# Chapter 5 Understanding Self-Renewal and Early Differentiation in Cancer Stem Cells

## 5.1 Summary

Similarities between stem cells and cancer cells led to the concept of the existence of cancer stem cells. Stem cells are defined by their ability to self-renew and to generate differentiated progeny, properties shared by cancer stem cells. The aim of this study was to establish gene expression profiles of self renewal and differentiation in normal stem cells as well as cancer stem cells thus allowing the study of genes required to drive cancer cell proliferation and progression. To do this, expression array analysis was performed on murine embryonic stem cells (ES) in their undifferentiated and differentiated states and compared with murine pluripotent and nullipotent teratocarcinoma cells in both their undifferentiated and differentiated states, which represent the cancer stem cell. All microarrays were performed in triplicate using Applied Biosystems technology. Data was analysed using a combination of R and Spotfire software programs. Differentiated and undifferentiated array data was compared for each cell line to obtain a list of genes differentially expressed upon differentiation. Putative gene lists including Homeobox transcription factors and hedgehog signalling pathway molecules were compiled for those genes involved in self-renewal and early differentiation of normal and cancer stem cells. Validation of array data was performed using TaqMan® chemistry from Applied Biosystems. The proposal is that these lists represent transcriptome profiles of self-renewal and differentiation in cancer stem cells and outline important pathways in early differentiation in both the normal and malignant states thus providing targets for future cancer therapies.

## 5.2 Introduction

The concept of a cancer stem cell has been in existence for a long time, as far back as 1974 as proposed by Pierce (Pierce, 1974). Much of the initial work on the concept of cancer stem cells was documented in haematological malignancies such as AML and CML but has since been recognised in many tumours. The defining features of stem cells are their ability to self-renew and to generate differentiated progeny, properties shared by cancer stem cells. Teratocarcinomas are malignant tumours that occur in the gonads and are largely composed of

undifferentiated cells with pluripotent capabilities. These tumours are often referred to as the classical stem cell tumour.

## 5.3 Aims

The aim of this chapter was to establish gene expression profiles of self renewal and differentiation in normal stem cells as well as cancer stem cells using teratoma tumourigenesis as our model system thus allowing us to study genes that are required to drive cancer cell proliferation and progression. A secondary aim was to analyse the differences between the nullipotent and pluripotent cell, thus producing a transcriptome profile of nullipotency.

## 5.4 Materials and Methods

### 5.4.1 Sample collection

The following three cells lines; murine embryonic stem cell line, ES-E14TG2a, pluripotent teratocarcinoma cell line, SCC-PSA1 and the nullipotent teratocarcinoma cell line, Nulli-SCC1 were cultured as per protocol (see Chapter 2, section 2.1.1, 2.1.2, 2.1.3). Cells were harvested at day 0 (1$^{st}$ passage) in triplicate. All cell lines were then allowed to spontaneously differentiate as per protocol for 3 days as well as maintaining some in their undifferentiated state for a subsequent passage. Cells were harvested following 3 days differentiation in triplicate. Cells were centrifuged and all spent medium was removed. Harvested cell samples were frozen at -80°C until RNA extraction was performed. Total number of samples was 27.

Samples were labelled as follows:

***E = ES-E14TG2a (murine embryonic stem cell line)***

a) E1 – E3: Samples in triplicate from day 0 (1$^{st}$ passage).
b) E4 – E6: Sample in triplicate from day 3, undifferentiated (2$^{nd}$ passage).
c) E7 – E9: Samples in triplicate from day 3, differentiated.

*S = SCC-PSA1 (pluripotent teratocarcinoma cell line)*

a) S1 – S3: Samples in triplicate from day 0 (1st passage).
b) S4 – S6: Sample in triplicate from day 3, undifferentiated (2nd passage).
c) S7 – S9: Samples in triplicate from day 3, differentiated.

*C = Nulli-SCC1 (nullipotent teratocarcinoma cell line)*

a) C1 – C3: Samples in triplicate from day 0 (1st passage).
b) C4 – C6: Sample in triplicate from day 3, undifferentiated (2nd passage).
c) C7 – C9: Samples in triplicate from day 3, differentiated.

### 5.4.2 RNA extraction and in vitro transcription (IVT)

RNA extraction and in vitro transcription were performed using Qiagen kits and Applied Biosystems technology as described in Chapter 2, section 2.3. The samples with their associated concentrations are listed in Table 5.2. Array hybridisation and reading of arrays was performed using Applied Biosystems technology as outlined in detail in Chapter 2, section 2.5.

### 5.4.3 Data analysis

The AB1700 package for R software (a free language and environment for statistical computing and graphics, R Development Core Team, 2004) was employed to filter data using a signal/noise ratio threshold > 3 in at least one sample. It read the output from the AB1700 software with normalisation of data and performance of *t*-test and fold change with graphics to visualise *t*-test results. Fold change values were calculated for each gene filtered. Data was imported into Spotfire® (Spotfire AB, Sweden) for further analysis. A number of comparisons were performed. Differentiated and undifferentiated array data was compared for each cell line to generate a list of differentially expressed genes responsible for differentiation in both states. Only genes with a fold change ≥ 2 and a false discovery rate (FDR) of ≤ 0.05 were considered (Reiner et al., 2003). The gene ontology site, PANTHER (**P**rotein **AN**alysis **TH**rough

Evolutionary Relationships), was used to identify the pathways, biological processes and molecular functions associated with significant genes. Hierarchical clustering of data was performed using Cluster 3.0 and Treeview 1.04 as described in Chapter 2, section 2.5.4.4 and 2.5.4.5.

**Table 5.1: Groups used for analysis of microarray data**

| Analysis No. | Class 1 | Class 2 |
|---|---|---|
| 1 | Undifferentiated ES cells (Day 0) | Differentiated ES cells (Day 3) |
| 2 | Undifferentiated SCC-PSA1 cells (Day 0) | Differentiated SCC-PSA1 cells (Day 3) |
| 3 | 'Undifferentiated' Nulli-SCC1 cells (Day 0) | 'Differentiated' Nulli-SCC1 cells (Day 3) |
| 4 | Analysis number 1 | Analysis number 2 |
| 5 | Undifferentiated SCC-PSA1 cells | 'Undifferentiated' Nulli-SCC1 cells |

### 5.4.4 Quantitative Real-time PCR analysis

Two μg of total RNA from each of the cell line samples was converted to cDNA using the High Capacity cDNA Archive kit (Applied Biosystems) following the manufacturer's protocol. The cDNA was used for quantitative real-time PCR amplification with TaqMan® chemistry (Applied Biosystems) using 47 pre-designed TaqMan® Gene Expression Assays from Applied Biosystems. Values were normalised relative to mouse GAPDH.

## 5.5 Results

### 5.5.1 RNA analysis

Quality and quantity of 18 total RNA samples was assessed by gel electrophoresis, spectrophotometry and using the Agilent 2100 Bioanalyser prior to use in expression array experiments. Concentrations of the total extracted RNA and amount of RNA used in RT-IVT reactions are shown in Table 5.2.

**Table 5.2: Concentrations (ng/µl) for samples used for Expression array analysis.**

| Sample | | Total RNA Conc. (µg/µl) | Amount for RT-IVT (µg) | Sample | | Total RNA Conc. (µg/µl) | Amount for RT-IVT (µg) |
|---|---|---|---|---|---|---|---|
| No | ID | | | No | ID | | |
| 1 | E1 | 0.710 | 2 | 10 | S7 | 2.210 | 2 |
| 2 | E2 | 0.870 | 2 | 11 | S8 | 3.602 | 2 |
| 3 | E3 | 0.492 | 2 | 12 | S9 | 2.192 | 2 |
| 4 | E7 | 1.816 | 2 | 13 | C1 | 1.092 | 2 |
| 5 | E8 | 2.012 | 2 | 14 | C2 | 1.152 | 2 |
| 6 | E9 | 2.726 | 2 | 15 | C3 | 0.568 | 2 |
| 7 | S1 | 0.568 | 2 | 16 | C7 | 1.386 | 2 |
| 8 | S2 | 1.644 | 2 | 17 | C8 | 0.308 | 2 |
| 9 | S3 | 1.138 | 2 | 18 | C9 | 0.204 | 2 |

### 5.5.2 cRNA analysis post RT-IVT

Post RT-IVT, cRNA was assessed qualitatively and quantitatively prior to array hybridisation. An example of a typical sample electrophoresed on an ethidium bromide-stained agarose gel is shown in Figure 5.1.

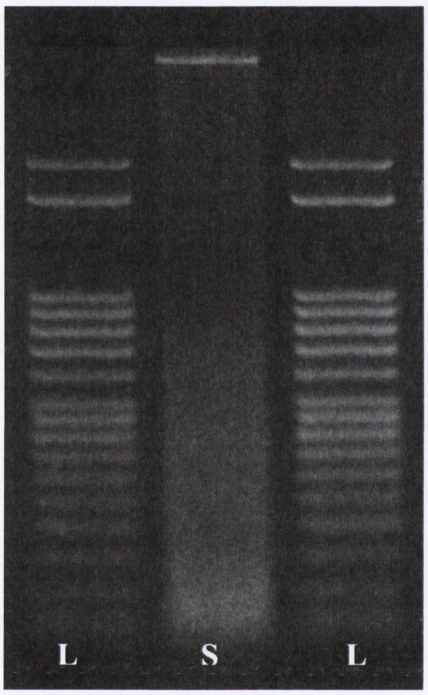



**Figure 5.1: Agarose gel electrophoresis of a cRNA sample. L = ladder (size marker), S = cRNA sample.**

cRNA concentration was also assessed prior to array hybridisation with 10μg of cRNA used per array (see Table 5.3 for details).

**Table 5.3: Samples and their total RNA concentrations, cRNA concentrations and amount of cRNA used for array hybridisation.**

| Sample | Total RNA conc. (µg/µl) | cRNA conc. (µg/µl) | Amount for Array hybridisation (µg) |
|---|---|---|---|
| **E1** | 0.71 | 0.1788 | 10 |
| **E2** | 0.87 | 0.1884 | 10 |
| **E3** | 0.492 | 0.4584 | 10 |
| **E7** | 1.816 | 02148 | 10 |
| **E8** | 2.012 | 0.2700 | 10 |
| **E9** | 2.726 | 0.3756 | 10 |
| **S1** | 0.568 | 0.1980 | 10 |
| **S2** | 1.644 | 0.4368 | 10 |
| **S3** | 1.138 | 0.2892 | 10 |
| **S7** | 2.21 | 0.3192 | 10 |
| **S8** | 3.602 | 0.2940 | 10 |
| **S9** | 2.192 | 0.3780 | 10 |
| **C1** | 1.092 | 0.3300 | 10 |
| **C2** | 1.152 | 0.3012 | 10 |
| **C3** | 0.568 | 0.3756 | 10 |
| **C7** | 1.386 | 0.5160 | 10 |
| **C8** | 0.308 | 0.6444 | 10 |
| **C9** | 0.204 | 0.5688 | 10 |

### 5.5.3 Statistical analysis

#### *5.5.3.1 Normalisation*

A 5% trimmed mean was the main normalisation method used in this study; this method produces reliable results when analysing low signal intensity genes. Figure 5.2 shows MA plots after normalisation for the biological replicates E1 – E3, E7 – E9, S1 – S3, E7 – E9, C1 – C3 and C7 – C9. MA plots for replicates are symmetrical around zero and show no appreciable difference. The three replicates from each passage show closer similarities than a comparison of a replicate from each of the two groups, i.e. undifferentiated (day 0) and differentiated (day 3) cells. This would be as expected. The remaining arrays (E4-6, S4-6 and C4-6) were similar to their day 0 undifferentiated counterparts.

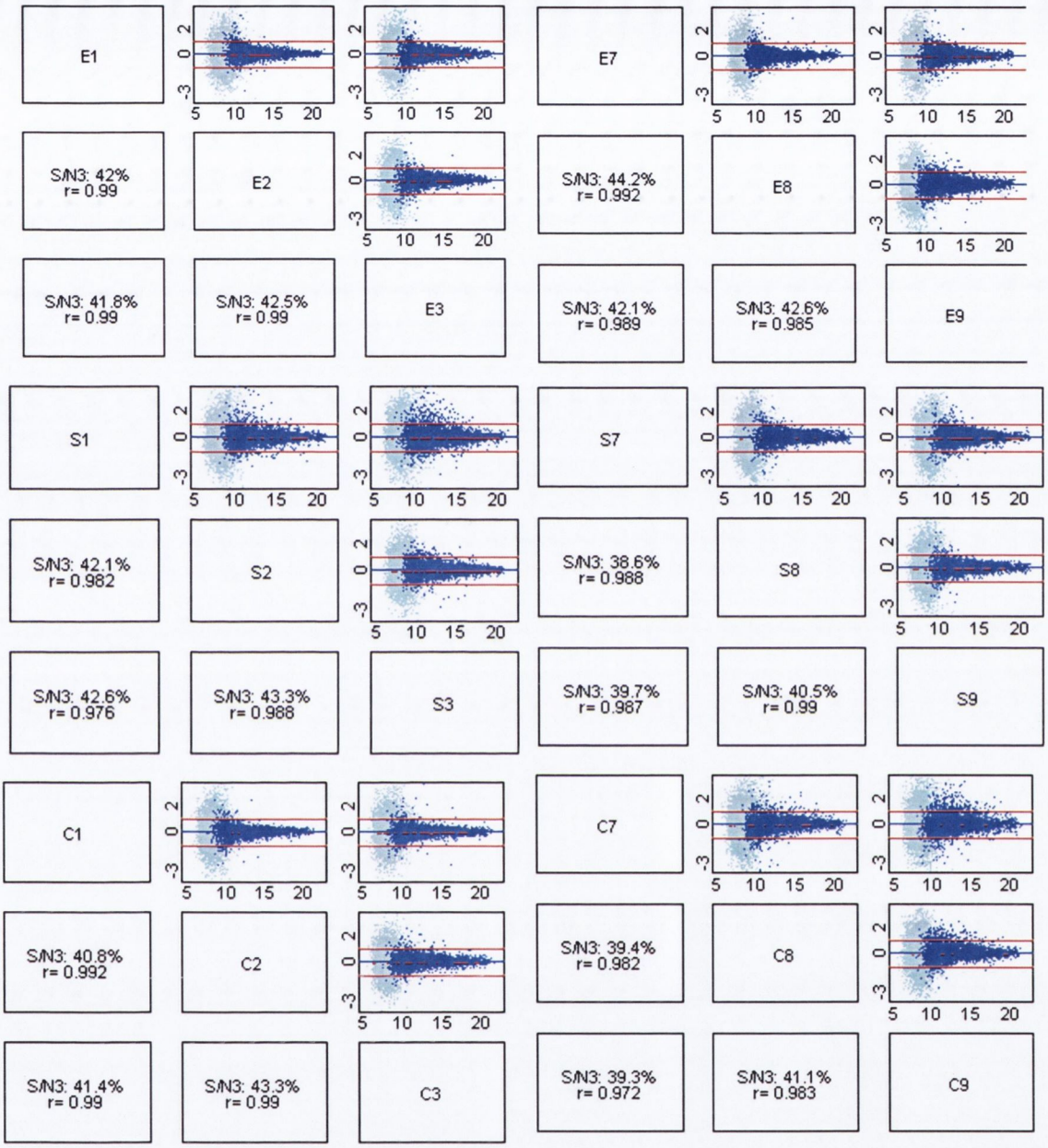



**Figure 5.2: MA plots of undifferentiated arrays (E/S/C1-3) and differentiated arrays (E/S/C7-9).**

These MA plots show good correlation between replicates in the undifferentiated and differentiated groups with r values of 0.99 in E1-3 and 0.985 – 0.992 in E7-9. The findings are similar in both the S and C cell lines.

#### *5.5.3.2 Filtering genes*

Step-wise filtering of gene probes was performed for all analyses as follows:

1) During normalisation in ab1700gui, probes with a signal-to-noise ratio of less than 3 (S/N < 3) in all arrays were eliminated.

2) Detectable genes were then analysed using a t test in ab1700gui and a correction for multiple testing. Genes with an FDR $\leq$ 0.05 were considered to show significant differential expression between two classes of analysis. Genes that were not significantly differentially expressed after t test or adjustment for multiple comparisons were eliminated.

3) Genes less than two-fold up- or down-regulated (FC < 2) between the two classes were eliminated.

### *5.5.3.3 Analysis of Data Groups*

***Analysis 1: Comparison of differentiated E cells (day 3) and undifferentiated E cells (day 0)***

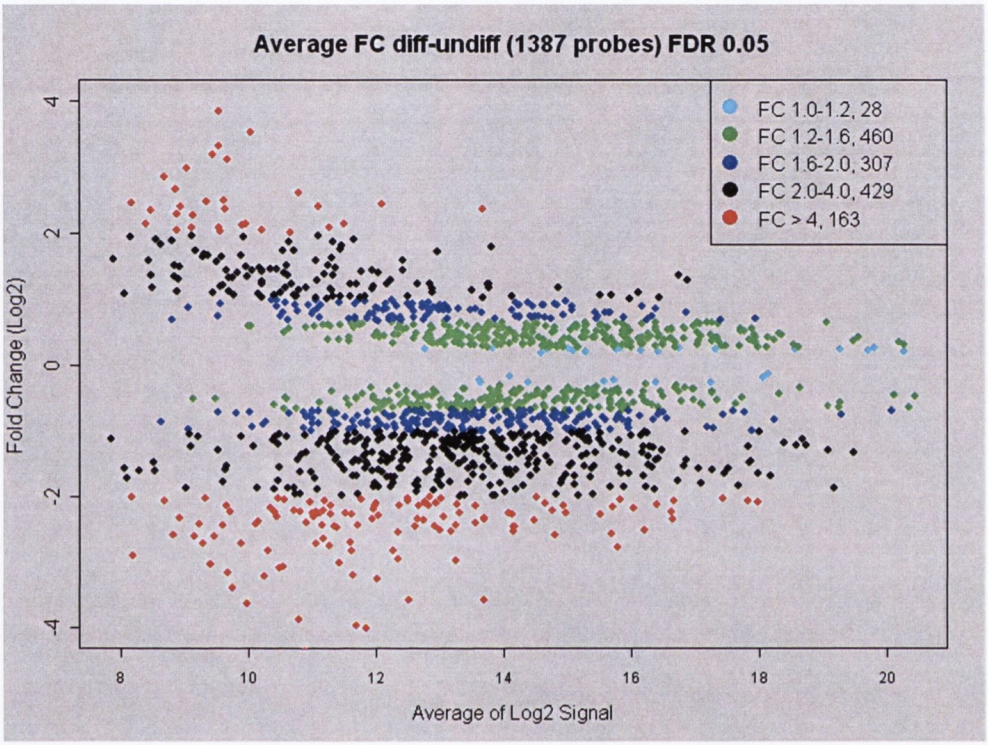



**Figure 5.3: Number of differentially expressed probes with FDR of 0.05 with their associated fold change values in a comparison of differentiated versus undifferentiated E cells.**

Number of gene probes with FDR ≤ 0.05= 1,387

- Genes with FC > 2 or < 0.5 = 592
- Number upregulated in differentiation = 149 (25.2%)
- Number downregulated in differentiation = 443 (74.8%)

**Table 5.4: Top 10 up and downregulated genes in differentiated v undifferentiated E cells**

| Down-regulated in diff E cells | | Up-regulated in diff E cells | |
|---|---|---|---|
| **Gene** | **Fold change** | **Gene** | **Fold change** |
| ***MP4*** | **0.014** | ***Mrgprh*** | **60.782** |
| ***Pck1*** | **0.020** | ***Trim7*** | **9.120** |
| ***Fgfrl1*** | **0.026** | ***Lemd1*** | **8.006** |
| ***1700047E16Rik*** | **0.049** | ***Bcl11b*** | **7.238** |
| ***Olfr508*** | **0.063** | ***E330034G19Rik*** | **6.407** |
| ***Eif5a2*** | **0.064** | ***Sema4d*** | **5.673** |
| ***9130015A21Rik*** | **0.069** | ***Kcnj3*** | **5.637** |
| ***Ccl5*** | **0.084** | ***Actc1*** | **5.291** |
| ***Oasl2*** | **0.084** | ***Helt*** | **4.903** |
| ***Trim30*** | **0.109** | ***Gpr132*** | **4.884** |

Using PANTHER, the pathways, biological processes and molecular functions that were over represented in E differentiated v undifferentiated gene list with FDR $\leq 0.05$ and FC $\geq 2$ and $< 0.5$ were identified.

**Table 5.5: PANTHER analysis of genes differentially expressed in the comparison of differentiated and undifferentiated E cells. The pathways in italics are those with a p value > 0.05 and thus considered not significantly over represented.**

| PANTHER Analysis of Differentially Expressed Genes | | |
|---|---|---|
| **Pathway** | **P value** | **Number of Genes** |
| Inflammation mediated by chemokine and cytokine signalling pathway | 0.003 | 19 |
| *Wnt signalling pathway* | *1* | *12* |
| *TGFβ signalling pathway* | *1* | *4* |
| *Integrin signalling pathway* | *1* | *11* |
| *Angiogenesis* | *1* | *6* |
| **Biological Process** | **P value** | **Number of Genes** |
| Tumour suppressor | 0.000285 | 11 |
| Interferon mediated immunity | 0.000865 | 10 |
| Developmental processes | 0.00289 | 68 |
| Oncogenesis | 0.00434 | 21 |
| Immunity and defense | 0.00954 | 49 |
| **Molecular function** | **P value** | **Number of Genes** |
| Extracellular matrix | 0.00000000274 | 30 |
| Extracellular matrix structural protein | 0.00000171 | 13 |
| Select calcium binding protein | 0.00000717 | 21 |
| Signalling molecules | 0.000107 | 36 |
| Extracellular matrix glycoprotein | 0.00069 | 11 |

*Analysis 2: Comparison of differentiated S cells (day 3) and undifferentiated S cells (day 0)*

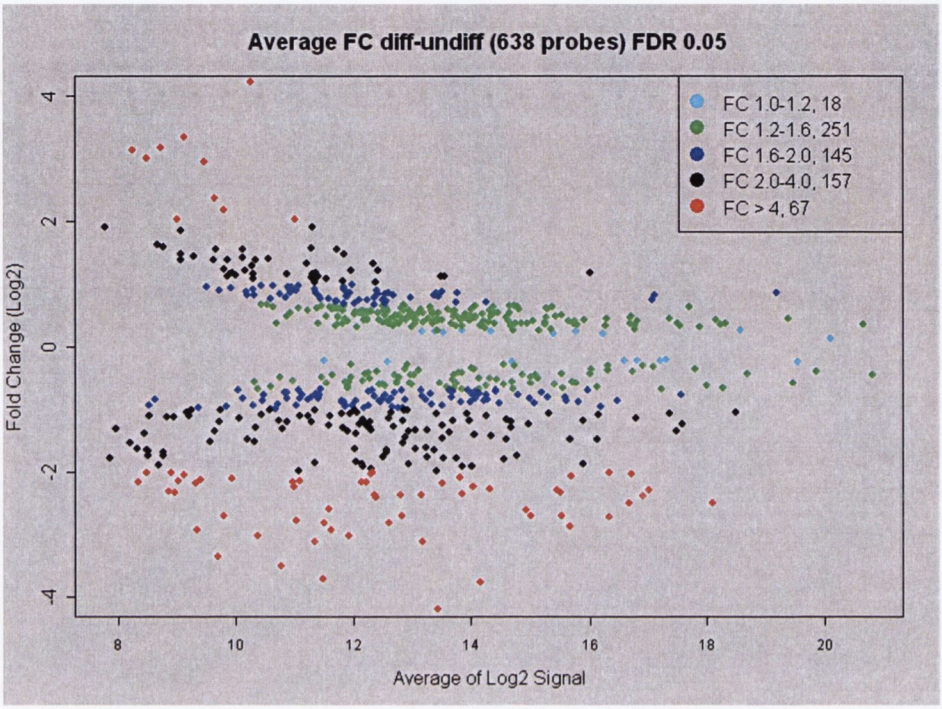



**Figure 5.4: Number of differentially expressed probes with FDR of 0.05 with their associated fold change values in a comparison of differentiated versus undifferentiated S cells.**

Number of gene probes with FDR ≤ 0.05=638

- Genes with FC ≥ 2 or < 0.5= 224
- Number upregulated upon differentiation = 56 (25%)
- Number downregulated upon differentiation (i.e. self renewal) = 168 (75%)

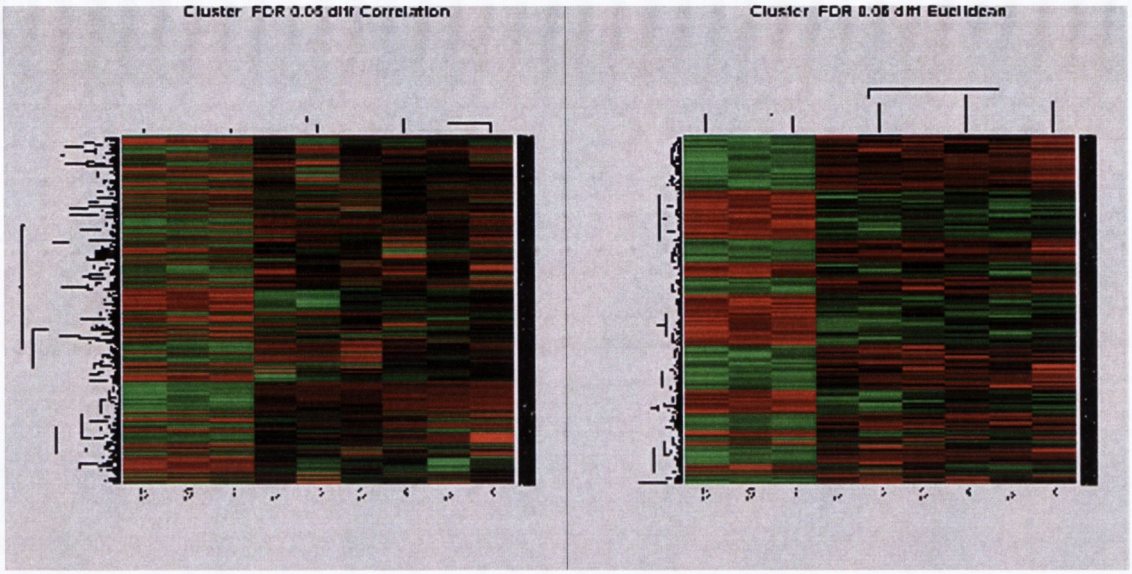



**Figure 5.5: Hierarchical clustering of undifferentiated and differentiated S cells.**

The undifferentiated S arrays (S1-6) are clustering together on the left hand side of these diagrams with the differentiated S arrays (S7-9) clustering tightly together on the right, as expected.

**Table 5.6: Top 10 up and downregulated genes in differentiated v undifferentiated S cells. Genes highlighted in yellow are also present in the same comparison in E cells.**

| Top 10 Differentially Expressed Genes in S cell line Differentiation | | | |
|---|---|---|---|
| **Down-regulated in diff. S cells** | | **Up-regulated in diff. S cells** | |
| **Gene** | **Fold change** | **Gene** | **Fold change** |
| *Ccl5* | 0.055 | *Olfr1450* | 8.868 |
| *Sdpr* | 0.077 | *Fgf5* | 7.747 |
| *Clec2d* | 0.089 | *Dscam* | 5.218 |
| *Gria1* | 0.133 | *Igfbp5* | 4.106 |
| *Ogn* | 0.133 | *Hist1h2ac* | 4.102 |
| *S100a4* | 0.137 | *Tmprss2* | 3.734 |
| *Ifi202b* | 0.143 | *Hsd17b9* | 3.604 |
| *Oasl2* | 0.144 | *1700027N10Rik* | 3.089 |
| *Tgfb3* | 0.146 | *Cthrc1* | 2.987 |
| *Dcn* | 0.152 | *Mal2* | 2.936 |

Using PANTHER, the pathways, biological processes and molecular functions that were over represented in S differentiated v undifferentiated gene list with FDR ≤ 0.05 and FC ≥ 2 and < 0.5 were identified.

**Table 5.7: PANTHER analysis of genes differentially expressed in the comparison of differentiated and undifferentiated S cells. The pathways in italics are those with a p value > 0.05 and thus considered not significantly over represented.**

| PANTHER Analysis of Differentially Expressed Genes | | |
|---|---|---|
| **Pathway** | **P value** | **Number of Genes** |
| Angiogenesis | 0.034 | 9 |
| *Inflammation mediated by chemokine and cytokine signalling pathway* | *0.076* | *10* |
| *VEGF signalling pathway* | *0.987* | *4* |
| *Endothelin signalling pathway* | *1* | *3* |
| *Wnt signalling pathway* | *1* | *6* |
| **Biological Process** | **P value** | **Number of Genes** |
| Cell proliferation and differentiation | $7.24 \times 10^{-7}$ | 27 |
| Immunity and defense | $2.66 \times 10^{-6}$ | 35 |
| Developmental processes | $9.22 \times 10^{-3}$ | 43 |
| Signal transduction | $1.47 \times 10^{-5}$ | 63 |
| Cell communication | $5.46 \times 10^{-3}$ | 25 |
| **Molecular function** | **P value** | **Number of Genes** |
| Signalling molecule | $2.51 \times 10^{-9}$ | 29 |
| Extracellular matrix | $1.42 \times 10^{-4}$ | 14 |
| Chemokine | $3.92 \times 10^{-3}$ | 5 |
| Annexin | $6.26 \times 10^{-3}$ | 6 |
| Other cytokine | $1.09 \times 10^{-2}$ | 4 |

*Analysis 3: Comparison of 'differentiated' C cells (day 3) and 'undifferentiated' Cells (day 3)*

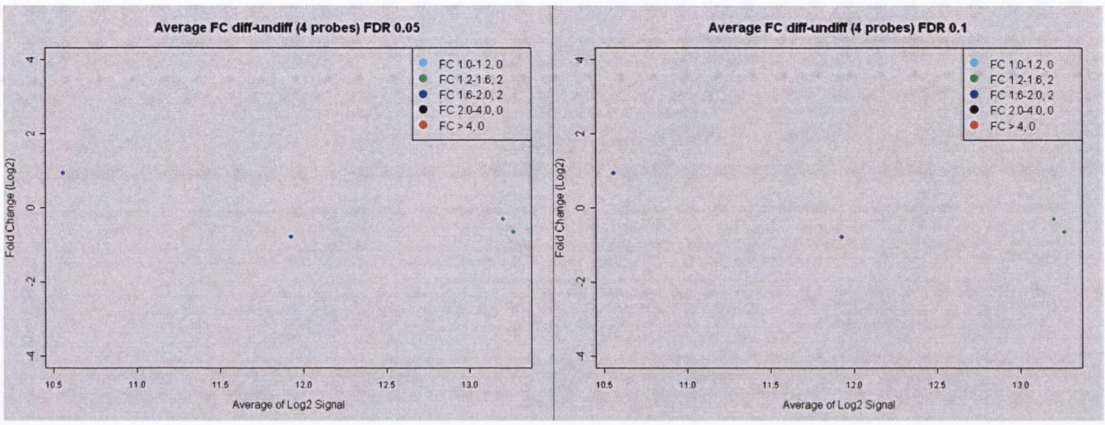



**Figure 5.6: Number of gene probes with FDR of 0.05 and 0.1 in C cell line.**

With a FDR ≤ 0.05, only 4 genes were differentially expressed between differentiated and undifferentiated C cells with all 4 probes having a fold change value < 2. The FDR was increased to 0.1 to see how many more genes would appear to be differentially expressed, but again only 4 were differentially expressed. As this cell line is nullipotent cell line and is not capable of differentiation, these results correlate with what was expected.

**Table 5.8: Genes differentially expressed between 'differentiated' and 'undifferentiated' Nullipotent teratocarcinoma cells.**

| | Gene Symbol | Gene Name | Fold change | P value |
|---|---|---|---|---|
| 1 | **Oat** | Ornithine aminotransferase | 0.63 | $4.99 \times 10^{-7}$ |
| 2 | **Unknown** | Unknown | 1.92 | $7.10 \times 10^{-6}$ |
| 3 | **Gnpat** | Glyceronephosphate O-acyltransferase | 0.58 | $6.75 \times 10^{-6}$ |
| 4 | **Abi2** | Abi-interacter 2 | 0.81 | $1.21 \times 10^{-5}$ |

*Analysis 4: Comparison of Analysis 1 and 2*

A fourth analysis was performed to compare the E and S cell lines to examine the differences and similarities between the differentially expressed genes upon differentiation in each group.

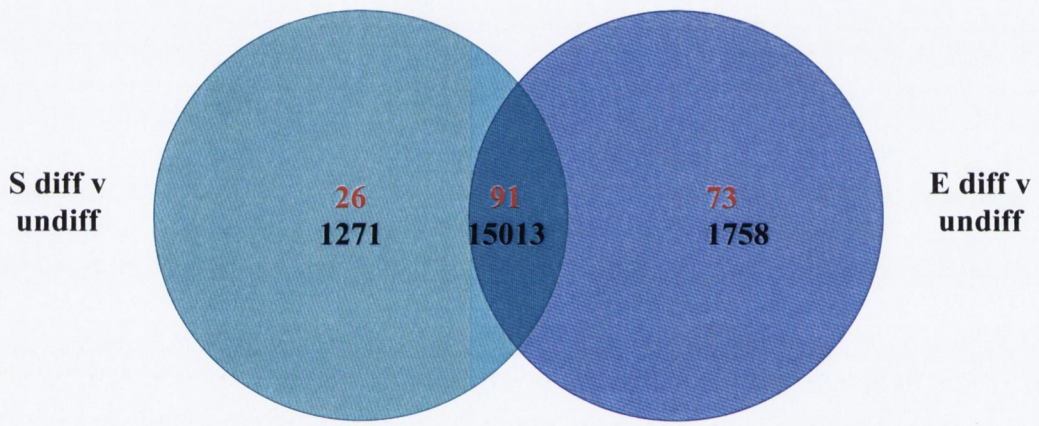



**Figure 5.7: Comparison of E and S cell lines. (Numbers in black = before removing genes with a fold change < 2 or FDR > 0.05, numbers in red = genes with a fold change ≥ 2 and FDR of ≤ 0.05).**

Analysis of the lists of differentially expressed genes imported to Spotfire® enabled the comparison of the data thus producing a list of what genes are changing upon differentiation in

S alone and E alone. It also indicates the genes that change upon differentiation that are common to S and E cells.

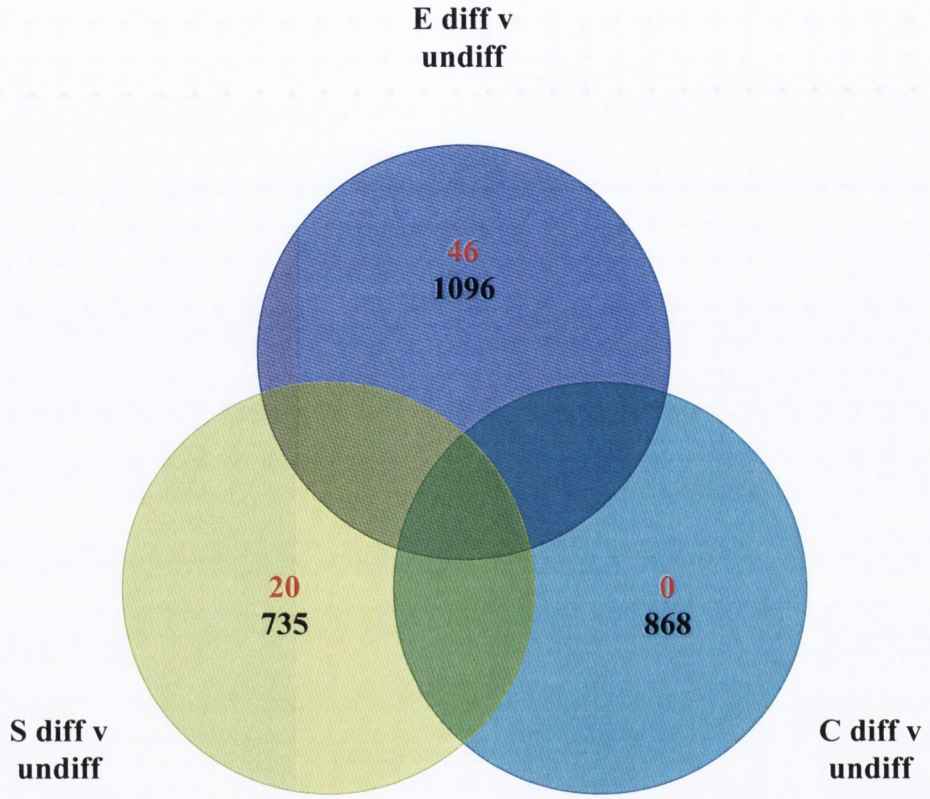



**Figure 5.8: Comparison of (E diff v undiff) v (S diff v undiff) v (C diff v undiff). (Numbers in black = before removing genes with a fold change < 2 or FDR > 0.05, numbers in red = genes with a fold change ≥ 2 and FDR of ≤ 0.05). Overlapping gene groups not illustrated.**

This gives a list of what genes are changing upon differentiation in E cell line alone, S cell line alone and C cell line alone.

**Table 5.9: 46 genes differentially regulated upon differentiation unique to E cell line (FDR < 0.05, FC > 2 or < 0.5). Blank spaces indicate those genes that are not known.**

| Differentially expressed genes upon differentiation unique to E cell line | | | |
|---|---|---|---|
| **Probe ID** | **FC (diff/undiff)** | **Gene Symbol** | **Gene Name** |
| 326533 | 0.134524627 | Hcn1 | hyperpolarization-activated, cyclic nucleotide-gated K+ 1 |
| 364848 | 0.151953493 | | |
| 911378 | 0.153624173 | | |
| 401142 | 0.190943584 | 4833427G06Rik | RIKEN cDNA 4833427G06 gene |
| 488138 | 0.228266086 | Sgcd | sarcoglycan, delta (dystrophin-associated glycoprotein) |
| 528710 | 0.23939909 | Sh2d1b | SH2 domain protein 1B |
| 807687 | 0.246176214 | | |
| 764190 | 0.249358911 | LOC238568 | similar to Serpinb6c protein |
| 698085 | 0.294493075 | Na | hypothetical LOC381739 |
| 520946 | 0.33671488 | | |
| 595928 | 0.361859165 | Tiam2 | T-cell lymphoma invasion and metastasis 2 |
| 479600 | 0.459270747 | Na | similar to novel protein |
| 516780 | 0.485873266 | Sstr2 | somatostatin receptor 2 |
| 632154 | 2.087749845 | B430119L13Rik | RIKEN cDNA B430119L13 gene |
| 816803 | 2.131810567 | | |
| 466930 | 2.22595329 | | |
| 912235 | 2.276254276 | Pcdh8 | protocadherin 8 |
| 342308 | 2.276306372 | Pogk | pogo transposable element with KRAB domain |
| 373586 | 2.345222899 | | |
| 468759 | 2.495546474 | | |
| 553600 | 2.535354424 | | |
| 829592 | 2.646224854 | Alox15 | arachidonate 15-lipoxygenase |
| 485721 | 2.708663355 | | |
| 606776 | 2.771136617 | 6430702L21Rik | RIKEN cDNA 6430702L21 gene |
| 663943 | 2.831166329 | Cmya4 | cardiomyopathy associated 4 |
| 323705 | 2.873336698 | | |
| 334779 | 3.053292878 | Faim2 | Fas apoptotic inhibitory molecule 2 |
| 746252 | 3.158801143 | Rab19 | RAB19, member RAS oncogene family |
| 397030 | 3.177926159 | 4933424A10Rik | RIKEN cDNA 4933424A10 gene |
| 846481 | 3.195426687 | 4432412L15Rik | RIKEN cDNA 4432412L15 gene |
| 913782 | 3.34232671 | A330019N05Rik | RIKEN cDNA A330019N05 gene |
| 790268 | 3.345633404 | Rxrg | retinoid X receptor gamma |
| 861162 | 3.754187459 | Wnt8a | wingless-related MMTV integration site 8A |
| 853192 | 3.834196313 | 5830467P10Rik | RIKEN cDNA 5830467P10 gene |
| 678570 | 3.863598919 | Grm6 | glutamate receptor, metabotropic 6 |
| 594302 | 3.886085609 | Na | similar to Protein C14orf115 |
| 404108 | 4.077951701 | Gm784 | gene model 784, (NCBI) |
| 705683 | 4.289946519 | 2810458H16Rik | RIKEN cDNA 2810458H16 gene |
| 493170 | 4.292215726 | Egr4 | early growth response 4 |
| 305846 | 4.397128232 | | |
| 889858 | 4.455312615 | | |
| 772538 | 5.291129908 | Actc1 | actin, alpha, cardiac |
| 627103 | 5.532756367 | | |
| 833252 | 5.673117398 | Sema4d | sema domain, immunoglobulin domain (Ig), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 4D |
| 780677 | 6.407344945 | E330034G19Rik | RIKEN cDNA E330034G19 gene |
| 658374 | 8.006448486 | Lemd1 | LEM domain containing 1 |

Table 5.9 shows the 46 genes unique to E cell differentiation with 13 (28.3%) downregulated upon differentiation and 33 (71.7%) upregulated upon differentiation. Many are unknown with 17 unnamed (37.0%).

**Table 5.10: 20 genes differentially expressed upon differentiation unique to S cell line differentiation.**

| Genes differentially expressed upon differentiation unique to S cell line | | | |
|---|---|---|---|
| **Probe ID** | **FC (diff/undiff)** | **Gene Symbol** | **Gene Name** |
| 429467 | 0.249377856 | Adamts5 | a disintegrin-like and metalloprotease |
| 688046 | 0.298666336 | | |
| 642099 | 0.322686288 | | |
| 461834 | 0.442078985 | Tnfrsf9 | tumor necrosis factor receptor superfamily, member 9 |
| 360072 | 2.183754996 | AU040377 | expressed sequence AU040377 |
| 612894 | 2.190166675 | Ugt2b34 | UDP glucuronosyltransferase 2 family, polypeptide B34 |
| 919185 | 2.232353982 | Tex12 | testis expressed gene 12 |
| 636855 | 2.261953875 | Sdsl | serine dehydratase-like |
| 666555 | 3.002802385 | | |
| 437440 | 3.603748026 | Hsd17b9 | hydroxysteroid (17-beta) dehydrogenase 9 |
| 396339 | 3.733561074 | Tmprss2 | transmembrane protease, serine 2 |
| 856242 | 3.770432457 | | |
| 729395 | 4.105995559 | Igfbp5 | insulin-like growth factor binding protein 5 |
| 312137 | 4.561419396 | | |
| 721264 | 5.217727561 | Dscam | Down syndrome cell adhesion molecule |
| 588033 | 8.073644148 | | |
| 652688 | 8.867792484 | Olfr1450 | olfactory receptor 1450 |
| 409154 | 9.028735918 | Na | similar to putative pheromone receptor |
| 485170 | 10.14260448 | | |
| 341289 | 18.60567886 | | |

Looking at the genes unique to S cell line differentiation in Table 5.10, 4 (20.0%) are downregulated and 16 (80.0%) are upregulated with almost half (9, 45%) being unnamed.

A comparison of PANTHER analysis obtained for both the E and S cell lines was also made and is described in the tables below.

**Table 5.11: Significantly over represented molecular functions in E diff v undiff and S diff v undiff taken from PANTHER. Common molecular functions are in italics.**

| E diff v undiff | | | S diff v undiff | | |
|---|---|---|---|---|---|
| **Molecular function** | **No** | **P value** | **Molecular function** | **No** | **P value** |
| *Extracellular matrix* | *30* | *0.00000000274* | *Signalling molecule* | *29* | *0.00000000251* |
| Extracellular matrix structural Protein | 13 | 0.00000171 | *Extracellular matrix* | *14* | *0.000142* |
| *Select calcium binding Protein* | *21* | *0.00000717* | Chemokine | 5 | 0.00392 |
| *Signalling molecules* | *36* | *0.000107* | Annexin | 6 | 0.00626 |
| Extracellular matrix glycoprotein | 11 | 0.00069 | Other cytokine | 4 | 0.0109 |
| Annexin | 9 | 0.00205 | *Select calcium binding proteins* | *9* | *0.0239* |
| Calmodulin related proteins | 13 | 0.00243 | Extracellular matrix glycoprotein | 6 | 0.0351 |
| | | | Defense/immunity protein | 12 | 0.00948 |

**Table 5.12: Significantly over represented biological processes in E diff v undiff and S diff v undiff taken from PANTHER. Common biological processes are in italics.**

| E diff v undiff | | | S diff v undiff | | |
|---|---|---|---|---|---|
| **Biological process** | **No** | **P value** | **Biological process** | **No** | **P value** |
| Tumour suppressor | 11 | 0.000285 | Cell proliferation and differentiation | 27 | 0.00000000724 |
| *Interferon mediated immunity* | *10* | *0.000865* | *Immunity and defense* | *35* | *0.00000266* |
| *Developmental processes* | *68* | *0.00289* | *Developmental processes* | *43* | *0.00000922* |
| *Oncogenesis* | *21* | *0.00434* | Signal transduction | 63 | 0.0000147 |
| *Immunity and defense* | 49 | 0.00954 | *Cell communication* | *25* | *0.00546* |
| *Cell communication* | *42* | *0.0228* | *Interferon-mediated immunity* | *6* | *0.0125* |
| Cell adhesion | 24 | 0.0307 | Oncogenesis | 6 | 0.0438 |
| Vision | 13 | 0.0416 | | | |

**Comparison of E and S cells:**

The majority (75%) of differentially expressed genes in both groups are downregulated upon differentiation. Ccl5 and Oasl2 are included in the top 10 differentiated expressed downregulated genes in both E and S cells. Ccl5 is a chemokine, a group of molecules overrepresented in the list of differentially expressed genes. Tumours that express Ccl5 have been shown in recent studies on lung and gastric cancers to progress and metastasise and to be associated with a poorer prognosis in general (Borczuk et al., 2008; Sugasawa et al., 2008). Oasl2 belongs to a family of interferon-induced antiviral proteins and have been found in humans and mice(Eskildsen et al., 2003). These genes are important components of the innate immunity in mammals. They have also shown to be involved in apoptosis, gene regulation , cell differentiation and growth (Perelygin et al., 2006).

Looking at the functional groups, within the biological processes, developmental processes and immunity and defense were common to both E and S cells and within molecular functions, signalling molecules and extracellular matrix functions were common to both groups.

**Signalling molecules:**

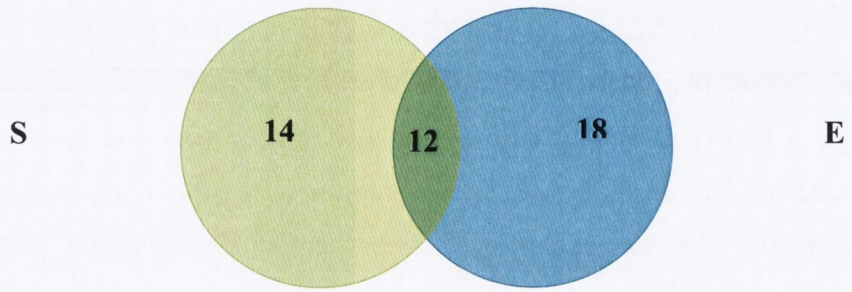



**Figure 5.9: Signalling molecules differentially expressed in E and S cell lines. 12 signalling molecules are differentially expressed upon differentiation in both E and S cell lines while 18 are unique to the E cell line and 14 are unique to the S cell line.**

**Table 5.13: Signalling molecules shared and unique to the E and S diff v undiff comparisons**

| | Signalling molecules | | |
|---|---|---|---|
| | **Shared** | **Unique to E** | **Unique to S** |
| **1** | Tgfb3 | Ppplrla | Cxcl1 |
| **2** | Cxcl10 | Traf5 | Pdgfc |
| **3** | Ccl5 | Cd37 | Sema7a |
| **4** | Ccl9 | Clec11a | Scgb1a1 |
| **5** | Ccl2 | Icam5 | Clec2d |
| **6** | Lsp1 | Ltb | Rho |
| **7** | Ltbp1 | Stc1 | Nrp1 |
| **8** | Spp1 (osteopontin) | S100a11 | Fbln5 |
| **9** | Fbln2 | Wnt8a | Fgf5 |
| **10** | Sema3b | Tspan11 | Mrpplf4 |
| **11** | Ltbp3 | Sema5a | Bdnf |
| **12** | s100a6 | Mgp | Crlf1 |
| **13** | | Elk3 | Plxnb3 |
| **14** | | Sema3a | Inhba |
| **15** | | Sh2d1b | |
| **16** | | Lemd1 | |
| **17** | | Sema4d | |
| **18** | | Tgfbi | |

Interestingly, chemokines are featuring highly in the shared group of signalling genes, namely Ccl2, 5 and 9. Other genes of note are Fbln2 (common to both groups) and Fbln5 (unique to S cell line).

***Analysis 5: Nullipotent (C cell line) versus Pluripotent (S cell line) teratocarcinoma cells***

For this analysis, the C cell line in its undifferentiated state was compared with the S cell line its undifferentiated state to look at the fundamental differences between the two cell lines.

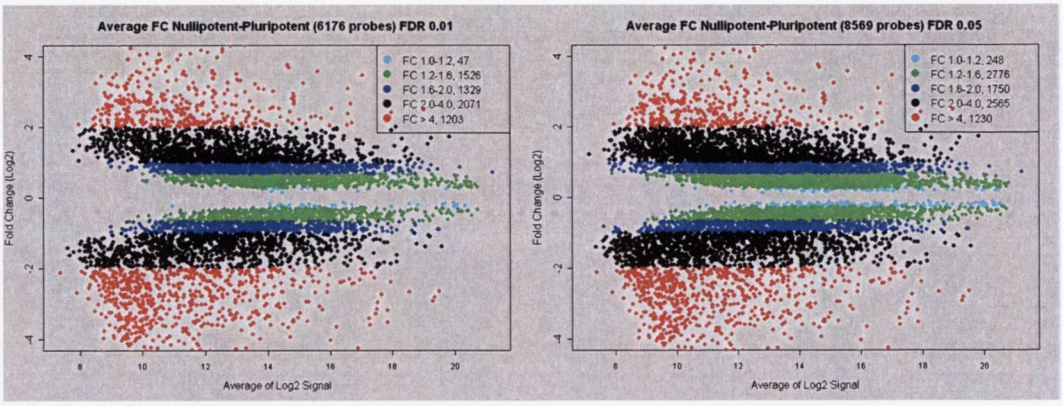



**Figure 5.10: Number of differentially expressed genes in C (nullipotent) v S (pluripotent) cells. FDR < 0.01 = 6,176 genes, FDR < 0.05 = 8,569.**

As the number of differentially expressed genes with a FDR < 0.05 was so high at 8,569, it was decided to be even more stringent with the data and use a FDR < 0.01 which produced 6,176 genes.

Considering only those with a fold change of > 2 or < 0.5:

- Genes with FC > 2 or < 0.5 = 3,275
- Number of upregulated genes in C cells = 1,536 (46.9%)
- Number of downregulated genes in C cells = 1,739 (53.1%)

**Table 5.14: Top 10 up and downregulated genes in undifferentiated C cell line v undifferentiated S cell line.**

| Top 10 up and downregulated genes | | | |
|---|---|---|---|
| **ProbeID** | **FC (Nulli/Plur)** | **Gene Symbol** | **Gene Name** |
| **Downregulated in C cell line compared with S cell line** | | | |
| 409948 | 0.001208367 | **Dcn** | Decorin |
| 338500 | 0.002713727 | **Hkdc1** | hexokinase domain containing 1 |
| 309042 | 0.003142657 | **Ctsh** | cathepsin H |
| 622680 | 0.003371231 | **Pthr1** | parathyroid hormone receptor 1 |
| 927537 | 0.00398058 | **Prg1** | proteoglycan 1, secretory granule |
| 343697 | 0.004181546 | **Tm4sf1** | transmembrane 4 superfamily member 1 |
| 483538 | 0.00428728 | **Fxyd3** | FXYD domain-containing ion transport regulator 3 |
| 930703 | 0.004554611 | **1810036H07Rik** | RIKEN cDNA 1810036H07 gene |
| 915023 | 0.004640483 | **Tnc** | tenascin C |
| 610341 | 0.00540789 | **Mglap** | matrix gamma-carboxyglutamate (gla) protein |
| **Upregulated in C cell line compared with S cell line** | | | |
| 590714 | 468.2714548 | **Pramel6** | Preferentially expressed antigen in melanoma like 6 |
| 861056 | 44.37380741 | **Rtn4rl1** | reticulon 4 receptor-like 1 |
| 821644 | 43.03644216 | **Hpgd** | hydroxyprostaglandin dehydrogenase 15 (NAD) |
| 744117 | 32.9144991 | **Rhbg** | Rhesus blood group-associated B glycoprotein |
| 824287 | 30.17320218 | **Crygd** | crystallin, gamma D |
| 351046 | 29.43998182 | **Clec4a1** | C-type lectin domain family 4, member a1 |
| 832438 | 29.24061827 | **Aard** | alanine and arginine rich domain containing protein |
| 732196 | 27.78586845 | **2610018G03Rik** | RIKEN cDNA 2610018G03 gene |
| 412244 | 27.4902195 | **Tm7sf4** | transmembrane 7 superfamily member 4 |
| 565267 | 27.002737 | **Tsx** | testis specific X-linked gene |

*PANTHER analysis:*

Using PANTHER to look at the pathways, biological processes and molecular functions that were over represented in the comparison of C v S cells produced the following:

**Table 5.15: PANTHER analysis of genes differentially expressed in the comparison of C v S cells. The pathways in italics are those with a p value > 0.05 and thus considered not significantly over represented.**

| PANTHER Analysis of Differentially Expressed Genes | | |
|---|---|---|
| **Pathway** | **P value** | **Number of Genes** |
| Angiogenesis | $2.16 \times 10^{-3}$ | 54 |
| *VEGF signalling pathway* | *0.147* | *22* |
| *Integrin signalling pathway* | *0.309* | *46* |
| *Notch signalling pathway* | *0.611* | *15* |
| *Alzheimer disease-presenilin pathway* | *0.732* | *29* |
| **Biological Process** | **P value** | **Number of Genes** |
| Developmental processes | $1.03 \times 10^{-8}$ | 374 |
| Oncogenesis | $3.38 \times 10^{-4}$ | 87 |
| Cell proliferation and differentiation | $1.24 \times 10^{-3}$ | 155 |
| Meiosis | $2.50 \times 10^{-3}$ | 27 |
| Cell structure and motility | $1.36 \times 10^{-2}$ | 170 |
| **Molecular function** | **P value** | **Number of Genes** |
| Signalling molecule | $4.76 \times 10^{-8}$ | 162 |
| Extracellular matrix | $9.65 \times 10^{-8}$ | 83 |
| Select regulatory molecule | $1.32 \times 10^{-4}$ | 203 |
| Growth factor | $7.19 \times 10^{-4}$ | 35 |
| Select calcium binding protein | $4.94 \times 10^{-3}$ | 58 |

**Table 5.16: Differential expression of the six differentiation markers used in Chapter 3 between C and S cells.**

| Gene Name | FC (C v S) | FDR | Probe ID |
|---|---|---|---|
| **Pou5f1** | 1.514 | 0.002 | 615439 |
| **Sox 2** | 1.529 | 0.001 | 848340 |
| **Nanog** | 3.201 | 1.25x10-6 | 593412 |
| **Afp** | 0.32 | 0.043 | 929310 |
| ***Ncam1*** | *1.005* | *0.987* | *772307* |
| ***Vegfr2*** | - | - | - |

Pou5f1 (Oct4), Sox2, Nanog and AFP all have FDR values < 0.05 with only Nanog and AFP having fold change values > 2 or < 0.5. Interestingly, Nanog is 3 fold upregulated in C cells compared with S cells.

### *5.5.4 Validation of Microarray Results*

Application of the 47 TaqMan® targets used in Chapter 4 and analysis in a relative quantitation study produced the following fold changes in a comparison of the differentiated state to the undifferentiated state in both E and S cell lines.

**Table 5.17: 47 targets validated in the differentiated v undifferentiated comparisons in the ES and Pluripotent cell lines with their values from the microarray and TaqMan® (TQ) experiments.**

**(Note: Numbers in italics refer to figures with insignificant p values and FDR values.)**

| | | E diff v undiff | | | S diff v undiff | | |
|---|---|---|---|---|---|---|---|
| | | TQ FC | Array FC | FDR | TQ FC | Array FC | FDR |
| 1 | **Adamts5** | 0.470 | Not found | | 0.142 | 0.249 | 0.018 |
| 2 | **AFP** | 0.234 | Not found | | 22.442 | *5.010* | *0.131* |
| 3 | **Ccnd1** | 0.253 | *0.459* | *0.054* | 0.200 | 0.246 | 0.014 |
| 4 | **Cdkn1c** | 2.409 | 2.162 | 0.011 | 1.351 | *1.002* | *0.997* |
| 5 | **Cited1** | 0.720 | *0.946* | *0.864* | 2.427 | 2.177 | *0.202* |
| 6 | **Dcn** | 0.093 | 0.240 | 0.018 | 0.212 | 0.152 | 0.010 |
| 7 | **Dscam** | 1.653 | Not found | | 4.544 | 5.218 | 0.037 |
| 8 | **Egr4** | 1.193 | 4.292 | 0.028 | 0.980 | Not found | |
| 9 | **Eif2s3y** | 0.853 | *1.000* | *0.999* | Undet in S | Not found | |
| 10 | **Fgf5** | 1.284 | *1.832* | *0.069* | 7.469 | 7.747 | 0.009 |
| 11 | **Foxj1** | 1.173 | 1.307 | 0.050 | 2.248 | *1.670* | *0.130* |
| 12 | **Fzd1** | 0.120 | 0.281 | 0.035 | 0.695 | *0.665* | *0.109* |
| 13 | **Galgt2** | 2.983 | *0.966* | *0.933* | 2.286 | *1.620* | *0.500* |
| 14 | **Gata1** | 0.594 | Not found | | 1.072 | Not found | |
| 15 | **Gata6** | 1.676 | 2.146 | 0.029 | 2.217 | *1.548* | *0.190* |
| 16 | **Gdf15** | 1.634 | 1.665 | 0.033 | 2.317 | 1.487 | 0.023 |
| 17 | **Hhip** | 0.588 | 1.558 | *0.231* | 1.915 | *1.017* | *0.994* |
| 18 | **Hoxb1** | 1.585 | *1.425* | *0.448* | 0.565 | Not found | |
| 19 | **Hoxd1** | 2.310 | Not found | | 1.280 | Not found | |
| 20 | **Hoxd9** | 1.779 | Not found | | 2.661 | *1.437* | *0.384* |
| 21 | **Igfbp5** | 2.988 | Not found | | 4.187 | 4.106 | 0.017 |
| 22 | **Klf2** | 1.227 | *2.306* | *0.104* | 1.523 | *1.102* | *0.803* |
| 23 | **Klk6** | 1.397 | *1.505* | *0.535* | 2.238 | *1.413* | *0.286* |
| 24 | **Lefty1** | 1.556 | *1.965* | *0.096* | 2.792 | *1.435* | *0.722* |
| 25 | **Lifr** | 1.356 | Not found | | 1.158 | *1.021* | *0.966* |
| 26 | **Mafb** | 0.819 | *1.136* | *0.817* | 1.491 | Not found | |
| 27 | **Mmp2** | 0.342 | *0.615* | *0.178* | 0.954 | *0.419* | *0.082* |
| 28 | **Myc** | 0.331 | *0.952* | *0.928* | 0.451 | *0.339* | *0.152* |
| 29 | **Nanog** | 0.520 | *1.053* | *0.797* | 1.803 | *1.141* | *0.447* |
| 30 | **Notch4** | 1.417 | *0.894* | *0.395* | 1.217 | *1.121* | *0.661* |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 31 | **Oct** | 0.685 | *1.283* | *0.057* | 1.278 | *0.950* | *0.588* |
| 32 | **Olfr1450** | Undet in E7 | Not found | | Undet in S | 8.868 | 0.018 |
| 33 | **Olig3** | 2.735 | *2.514* | *0.111* | 0.523 | Not found | |
| 34 | **Pramel6** | 0.856 | *1.010* | *0.979* | 0.376 | Not found | |
| 35 | **Rab15** | 1.218 | *1.714* | *0.121* | 1.686 | *1.258* | *0.526* |
| 36 | **Rhoj** | 0.158 | 0.327 | 0.045 | 0.256 | *0.147* | *0.130* |
| 37 | **Scml2** | 0.860 | *1.769* | *0.514* | 2.063 | *2.144* | *0.384* |
| 38 | **Slamf9** | 0.432 | Not found | | 1.402 | Not found | |
| 39 | **Sox17** | 1.602 | *1.453* | *0.206* | 2.009 | *1.186* | *0.437* |
| 40 | **Sox2** | 0.798 | 1.549 | 0.037 | 1.278 | *1.149* | *0.598* |
| 41 | **Tcstv1** | 0.206 | *1.400* | *0.345* | 0.885 | *0.884* | *0.866* |
| 42 | **Tert** | 1.629 | Not found | | 1.599 | *0.903* | *0.802* |
| 43 | **Tex12** | 1.527 | Not found | | 2.372 | 2.232 | 0.009 |
| 44 | **Timp2** | 0.177 | 0.290 | 0.019 | 0.972 | 0.455 | 0.024 |
| 45 | **Tnc** | 0.083 | 0.214 | 0.041 | 0.135 | *0.104* | *0.072* |
| 46 | **Tnfrsf9** | 0.145 | Not found | | 0.233 | 0.442 | 0.044 |
| 47 | **Wnt6** | 0.369 | *0.982* | *0.953* | 0.800 | *0.914* | *0.684* |

**Differentiation in mouse ES cells:**

Looking at only those validation targets with significant microarray fold changes results in only 11 out of the 47 targets being differentially expressed between the differentiated E and undifferentiated E cell group. As is demonstrated in Table 5.18, fold changes are very similar between TaqMan® and microarray with Figure 5.11 showing a correlation coefficient of 0.94.

**Table 5.18: 11 of the 48 genes were present in both TaqMan® and microarray experiments for normal differentiation. All microarray fold changes were significant (i.e. FDR < 0.05).**

| | Gene Symbol | E diff v undiff | |
|---|---|---|---|
| | | TQ | Array |
| 1 | Cdkn1c | 2.409 | 2.162 |
| 2 | Dcn | 0.093 | 0.240 |
| 3 | Egr4 | 1.193 | 4.292 |
| 4 | Foxj1 | 1.173 | 1.307 |
| 5 | Fzd1 | 0.120 | 0.281 |
| 6 | Gata6 | 1.676 | 2.146 |
| 7 | Gdf15 | 1.634 | 1.665 |
| 8 | Hhip | 0.588 | 1.558 |
| 9 | Rhoj | 0.158 | 0.327 |
| 10 | Timp2 | 0.177 | 0.290 |
| 11 | Tnc | 0.083 | 0.214 |

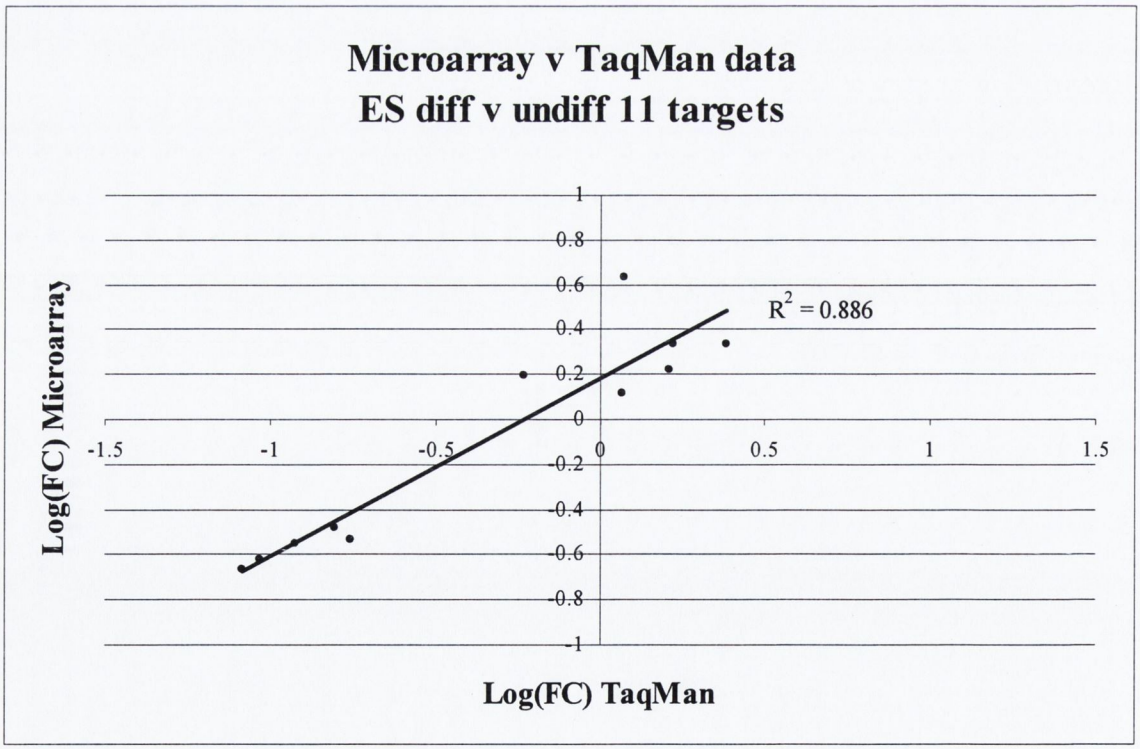



**Figure 5.11: Graph of Log (FC) TaqMan® v Log (FC) Microarray in E cell differentiation. (Correlation coefficient = 0.94, Slope = 0.79).**

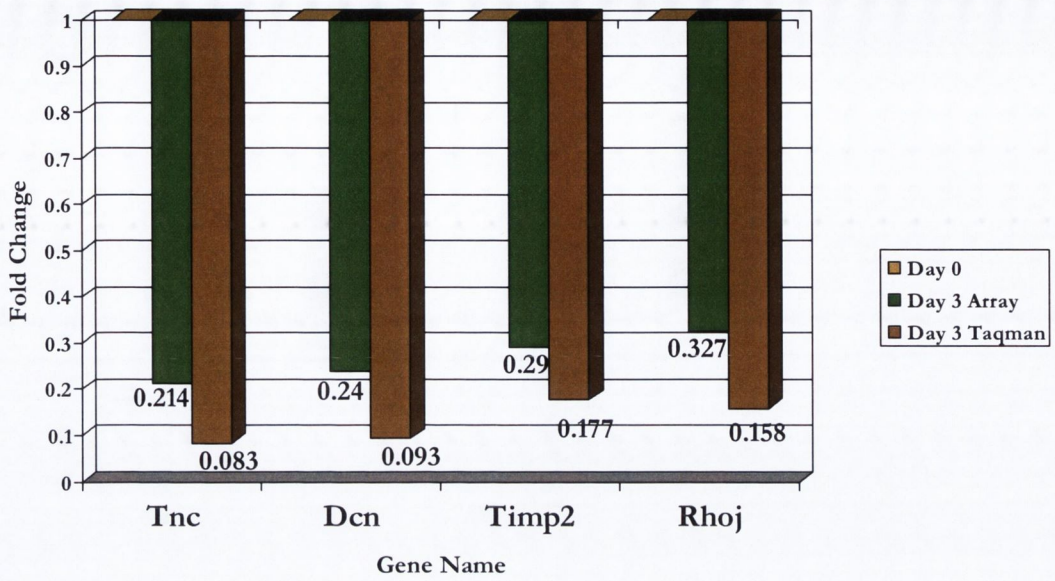



**Figure 5.12: Microarray v TaqMan® mRNA fold change. Genes downregulated in E cell line differentiation compared with undifferentiation.**

Figure 5.12 highlights a comparison of four downregulated genes in differentiated E cells compared with undifferentiated cells with Tnc (Tenascin C) and Dcn (Decorin) being two of the most downregulated genes in differentiated E cells. Results compare very well between the two methodologies of gene expression measurement.

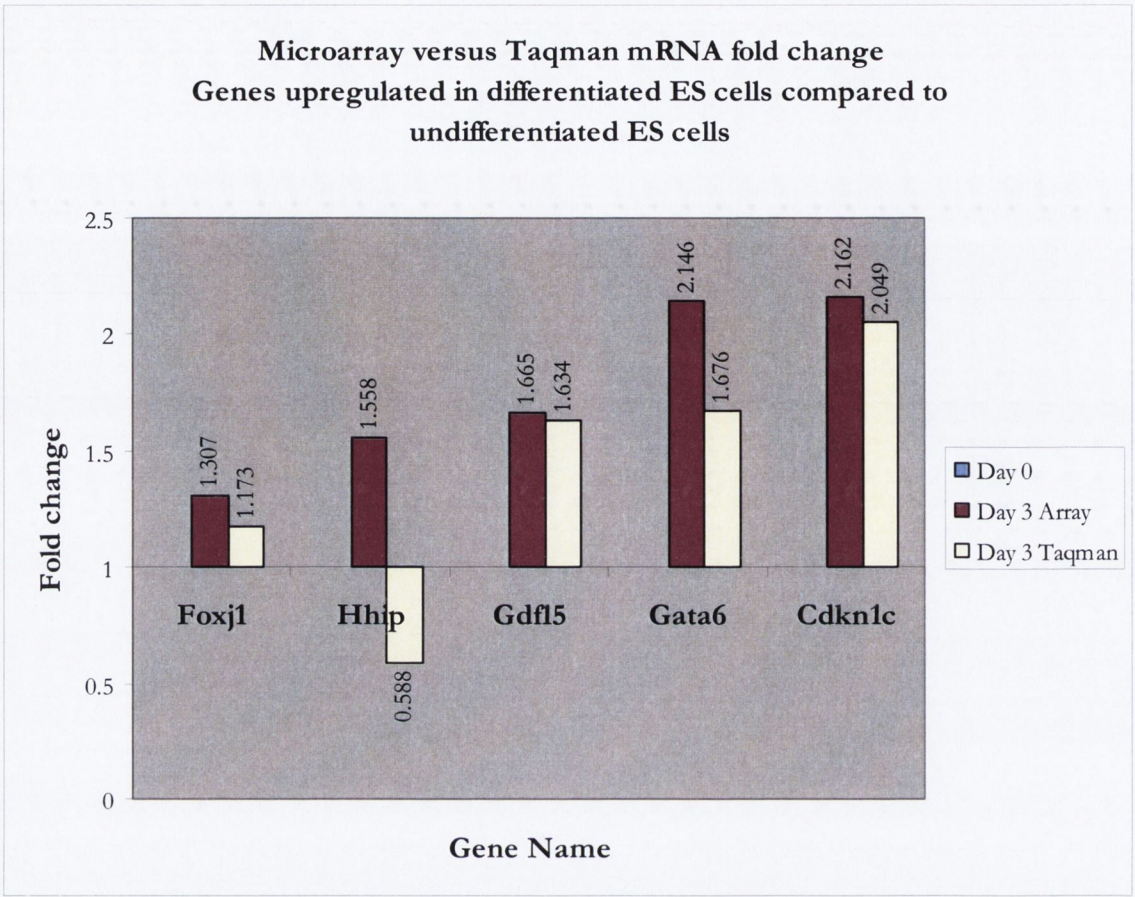



**Figure 5.13: Microarray v TaqMan® mRNA fold change. Genes upregulated in differentiated E cells compared with undifferentiated E cells.**

Figure 5.13 shows a selection of genes upregulated in differentiated E cells, again showing a good correlation between TaqMan® and microarray techniques with the exception of Hhip which was marginally upregulated on the microarray (1.558 fold) but downregulated by TaqMan® analysis (1.700 fold). While the measurements differ, the fold change is less than 2 in both cases.

**Differentiation in pluripotent teratocarcinoma cells:**

**Table 5.19: 10 of the 48 genes were present in both TaqMan® and microarray experiments for differentiation in pluripotent teratocarcinoma cells (S cell line). All microarray fold changes were significant (i.e. FDR < 0.05).**

| | Gene Symbol | S diff v undiff | |
|---|---|---|---|
| | | TQ | Array |
| 1 | Adamts5 | 0.142 | 0.249 |
| 2 | AFP | 22.442 | 5.010 |
| 3 | Ccnd1 | 0.200 | 0.246 |
| 4 | Dcn | 0.212 | 0.152 |
| 5 | Dscam | 4.544 | 5.218 |
| 6 | Gdf15 | 2.317 | 1.487 |
| 7 | Igfbp5 | 4.187 | 4.106 |
| 8 | Mmp2 | 0.954 | 0.419 |
| 9 | Tex12 | 2.372 | 2.232 |
| 10 | Tnfrsf9 | 0.233 | 0.442 |

Of the 47 validation targets chosen, only 10 were significantly differentially regulated by microarray analysis in the S differentiation comparison with AFP demonstrating the highest differential expression in differentiated S cells in both methods. All 10 results correspond well between both methodologies with no gene showing contrasting results.

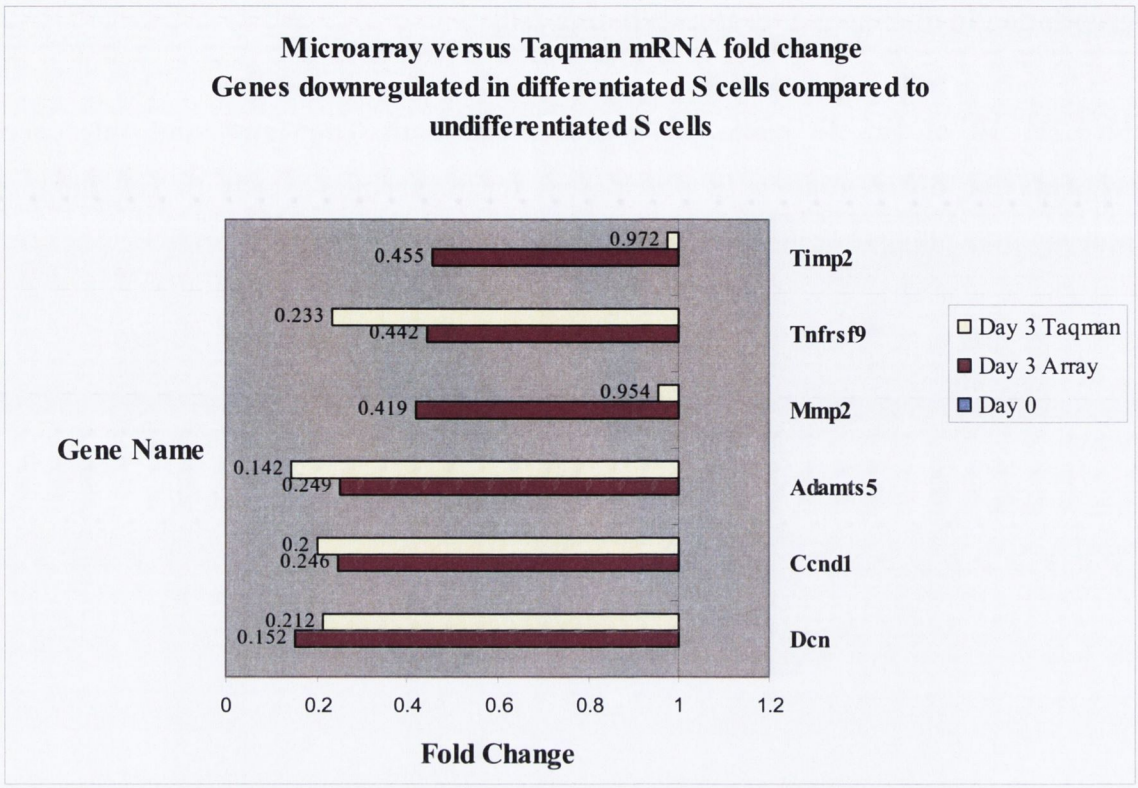



**Figure 5.14: Microarray v TaqMan® mRNA fold change. Genes downregulated in differentiated S cell compared with undifferentiated S cells.**

Downregulated genes, Dcn, Ccnd1, Adamts5, Mmp2, Tnfrsf9 and Timp2, are demonstrated in this graph with both gene expression measurement methods comparing well.

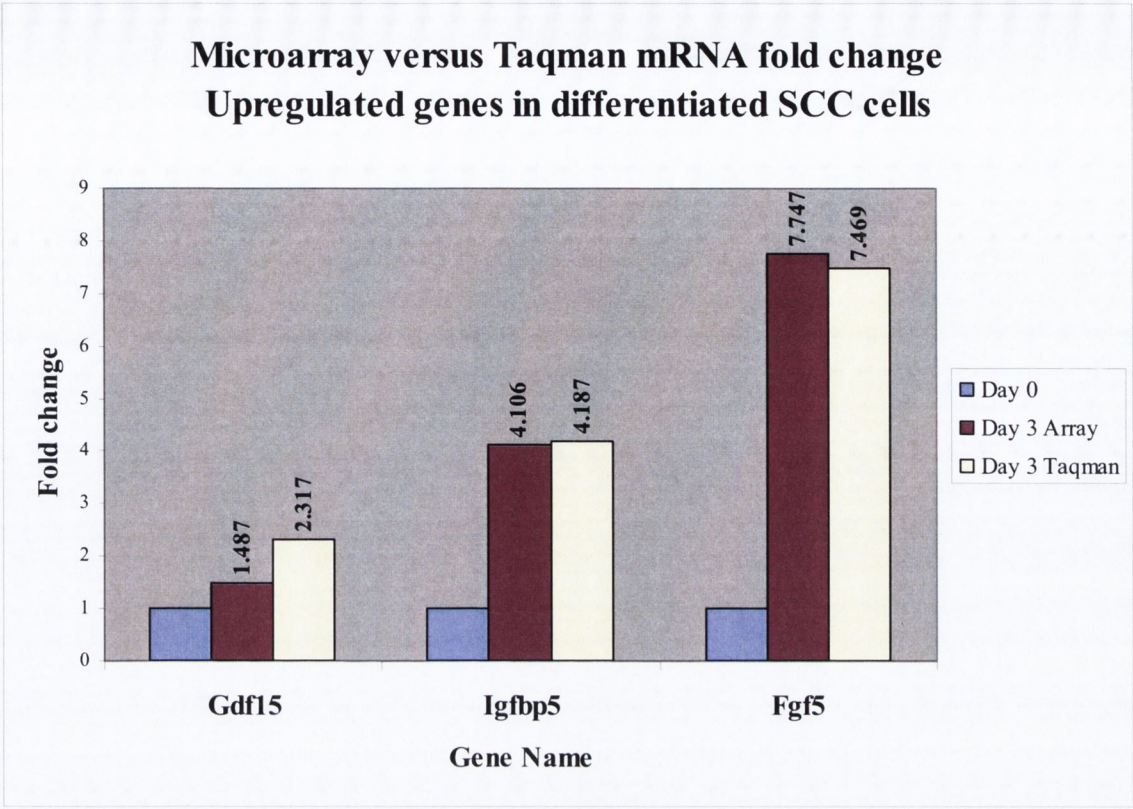



**Figure 5.15: Microarray v TaqMan® mRNA fold change. Upregulated genes in differentiated S cells.**

The upregulated genes in differentiated S cells are demonstrated in Figures 5.15 and 5.16, again with both methodologies comparing well. In general, the TaqMan® appears to give a higher fold change which has been recognised as TaqMan® is a more accurate and sensitive method of gene expression measurement.

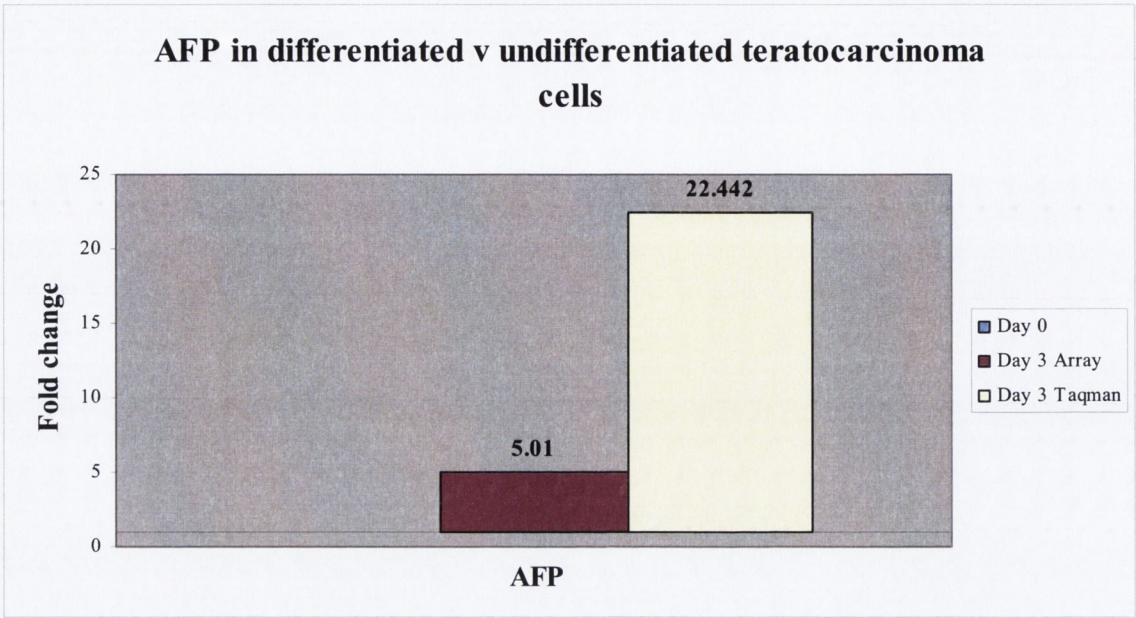



**Figure 5.16: AFP expression in differentiated S cells compared with undifferentiated using microarray and TaqMan® analysis.**

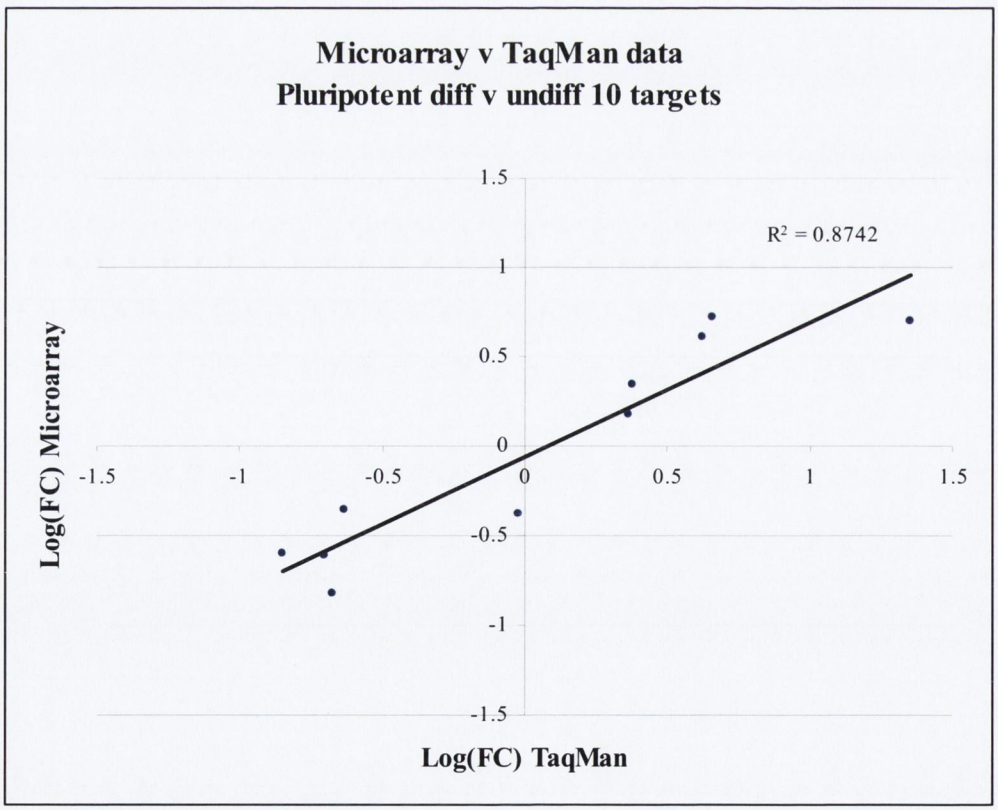



**Figure 5.17: Graph of Log (FC) TaqMan® v Log (FC) Microarray in pluripotent teratocarcinoma differentiation. (Correlation coefficient = 0.93, Slope = 0.75).**

This graph (Figure 5.17) shows an excellent correlation between TaqMan® and Microarray with a correlation coefficient of 0.93).

**Table 5.20: Comparison of common stem cell markers in 'benign' v 'malignant' stem cell.**

| | E diff v E undiff | | | S diff v S undiff | | |
|---|---|---|---|---|---|---|
| | TQ FC | Array FC | FDR | TQ FC | Array FC | FDR |
| **Nanog** | 0.520 | 1.053 | 0.797 | 1.803 | 1.141 | 0.447 |
| **Oct4** | 0.685 | 1.283 | 0.057 | 1.278 | 0.950 | 0.588 |
| **Sox2** | 0.798 | 1.549 | 0.037 | 1.278 | 1.149 | 0.598 |

Interestingly, the markers indicative of stem cells and the self renewing state, namely Nanog, Oct4 and Sox2, appeared to be slightly upregulated upon differentiation with similar results using both microarray and TaqMan® analysis. However, the FDR was not significant apart from Sox2 and the fold changes are small, being < 2 in all cases.

## 5.6 Discussion

Using cRNA microarrays with independent validation using real-time RT-PCR, this chapter presents a unique comparison of early differentiation in mES cells and their malignant counterpart, murine teratocarcinomas or embryonal carcinomas (specifically SCC-PSA1, a pluripotent teratocarcinoma cell line). Other authors have studied differentiation in mES cells but with a focus on differentiation at later time points and have not included differentiation or a comparison of same in a malignant cell line. The earliest time point examined by Heo et al, 2005 using gene expression array analysis was day 7 with further analysis at day 14, 21 and 28 (Heo et al., 2005). ***One of their most significant findings was that the majority of the transcriptional changes were occurring over the first 7 days suggesting that is an important time frame in which to examine the genes that turn on differentiation and thus one of the reasons for the focus on an early time point in this study***. They also found a predominance of downregulated genes upon differentiation similar to this study. However, unlike this study, their study does not include a comparison with differentiation in a malignant cell line.

Ccl5 and Oasl2 were two genes found to be significantly downregulated upon differentiation in both ES cells and their malignant counterpart. Ccl5 is a chemokine, a group of molecules significantly overrepresented as identified by PANTHER in the list of differentially expressed genes. Chemokines play an important role as soluble factors regulating migration of leucocytes during the inflammatory response, however, they also serve as potential sources of growth factors for tumour cells and angiogenic factors for endothelial cells (Kakinuma et al., 2006).

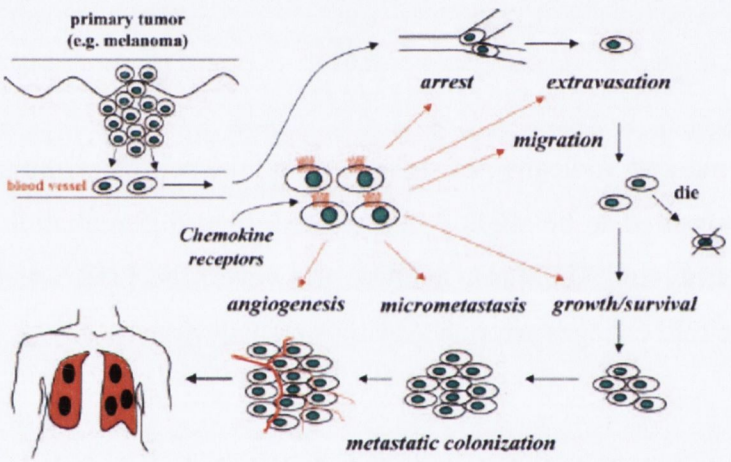



**Figure 5.18: Illustration of chemokines and their receptors playing a key role in several steps during the process of metastasis (Kakinuma et al., 2006).**

Indeed, there is increasing evidence that chemokines play an important role in promoting growth, survival, and metastasis of several malignancies (Gerard et al., 2001). Tumours that express Ccl5 have been shown in recent studies on ovarian, lung and gastric cancers to be more likely to progress and metastasise and are thus associated with a poorer prognosis than those tumours that lack Ccl5 expression (Borczuk et al., 2008; Sugasawa et al., 2008). It may be that expression of Ccl5 within a tumour indicates the presence of a population of cells maintaining their self renewal or stem cell-like properties thus permitting these tumours to progress and proliferate. Our data supports this idea with Ccl5 expression downregulated upon differentiation of teratocarcinoma stem cells as well as normal stem cells.

Oasl2 belongs to a family of interferon-induced antiviral proteins and have been found in both humans and mice (Eskildsen et al., 2003). These genes are important components of the innate immune system in mammals. They have been shown to be involved in apoptosis, gene regulation, cell differentiation and growth (Perelygin et al., 2006). There are no studies to date demonstrating Oasl2 expression in tumours. This study, however, suggests that it may have a role to play in maintaining self-renewal and its further characterisation may provide some useful information as we try to understand the mechanisms that determine and maintain the self-renewing state.

Transcription factors and extracellular matrix proteins are groups of genes found to be significantly over represented in our study, a similar finding to other studies on mES differentiation as well as human ES cells (Ivanova et al., 2002; Brandenberger et al., 2004b; Hailesellasse Sene et al., 2007). Signalling molecules are also significantly over represented in both our differentiation comparisons following analysis using PANTHER. As well as looking at common genes, the genes unique to mES cells and their malignant equivalent, pluripotent EC cells (SCC-PSA1) were extracted as they may provide an insight into the main differences between 'benign' and so called cancer stem cells. A danger with targeting cancer stem cells is the potential elimination of normal stem cells, a potentially significant side effect. Investigating genes unique to the cancer stem cell population would thus further improve the chances of a successful treatment. Fibulins comprise a small family of widely expressed extracellular matrix proteins that localise to basement membranes and stoma where they mediate cell-to-cell and cell-to-matrix communication. Fbln2 and 5 feature as significantly differentially expressed genes in this study with Fbln2 found to be downregulated upon differentiation (2.9 – ES cells and 5 fold – EC cells) in both sets of differentiation data and Fbln5 exclusively differentially regulated by the malignant cell line upon differentiation. Fbln2 has been shown in other studies to be downregulated upon differentiation in ES cells (Palmqvist et al., 2005). Reduced Fbln5 expression has been associated with tumour progression and proliferation in humans (Albig et al., 2006). It is thought to suppress tumour growth by inhibiting angiogenesis and has also been shown to inhibit epithelial cell proliferation with downregulation in the vast majority of epithelial-derived human tumours (Schiemann et al., 2002; Sullivan et al., 2007). The results in this chapter support the finding of reduced Fbln5 in tumour cells as Fbln5 expression is reduced in the malignant cell line upon differentiation and was not significantly differentially expressed in the ES cell line. This

has not been previously described. Therefore, this finding suggests that this gene as well as others found only within the malignant group may be targets for future cancer stem cell specific gene therapies that would have minimal effect on the normal stem cell populations whether they target the stem cell itself or its niche. The stem cell niche, defined as the complex microenviroment in which stem cells reside with other different cell types and the extracellular matrix molecules that dictate stem cell self-renewal and progeny production, is well recognised as a target for cancer therapies (Adams et al., 2007; Jones et al., 2008).

The genes governing nullipotency highlighted in this chapter in a comparison of nulli- and pluripotent embryonal carcinoma cells offer new insight into to the distinctive properties of the cancer stem cell. Pramel6 was the most upregulated gene in nullipotent cells. Preferentially expressed antigens of melanoma (Prame) are a group of tumour antigens shown to be over expressed in a wide variety of cancers and confer prognostic properties (Epping et al., 2006). They are known for their expression in melanomas with 88% of primary melanomas and 95% of metastatic melanomas expressed Prame antigens. PRAME expression in human solid tumours confers a poor clinical outcome with an increased likelihood of metastases and advanced stage contrasting with findings in AML where the presence of PRAME was associated with a good prognosis (Santamaria et al., 2008). A large family of PRAME-like genes and pseudogenes have evolved in the human genome with multiple PRAME-like genes also found in the mouse (Epping et al., 2006). Pramel6 has been described as being expressed in pre-implantation embryos and embryonic pluripotent stem cells in mice; however, its role in embryonic development has not been definitively elucidated. One suggestion is that in combination with other genes it has a role to play in determining the fate of ES cells (Kaji et al., 2006; Cinelli et al., 2008). Perhaps its over expression in nullipotent cells described in this chapter confers a survival advantage to these cells enabling them to remain in their undifferentiated self-renewing state. Cancer stem cells in possession of such a gene would therefore remain immune to conventional treatments and thus tumours demonstrating PRAME expression may have a poorer prognosis.

## 5.7 Conclusion

Within the cancer stem cell model, aberrant control of differentiation and self-renewal lie at the core of oncogenesis with dysregulation of self-renewal thought to be a key event in the development of cancer (Andrews et al., 2005; Clarke, 2005). Thus the study of genes that maintain pluripotency and those required for differentiation may ultimately be targets for cancer therapies in the future. This chapter presents a unique comparison of differentiation in mouse ES cells and EC cells and has identified a number of target genes not previously described in stem cells. Analysis of these cell lines produced a list of 20 genes unique to pluripotent EC cell differentiation, 2 of which, Adamts5 and Tnfsrf9, are downregulated upon differentiation. A list of 46 genes unique to mES cells was identified with 13 of these downregulated upon differentiation. These genes are proposed to represent the essential differences between benign and malignant stem cell differentiation and thus should be pursued to a more detailed functional level. This chapter also looked at the essential differences between the nullipotent and pluripotent EC cell producing a molecular signature of nullipotency composed of 1,536 genes upregulated in nullipotent compared with pluripotent EC cells.

# CHAPTER 6 SELF RENEWAL AND DIFFERENTIATION IN HUMAN EMBRYONAL CARCINOMA CELLS, PLURIPOTENT AND NULLIPOTENT

## 6.1 Summary

Similarities between stem cells and cancer cells have led to the concept of the existence of a population of cancer stem cells within a given tumour which is responsible for its continued growth and maintenance. The molecular mechanisms governing the regulation of the cancer stem cells are poorly understood and obtaining a better understanding would be a major advance in the field of tumour biology in terms of achieving more effective cancer treatments. The aim of this chapter was to establish gene expression profiles of differentiation and self-renewal in human pluri- and nullipotent embryonal carcinoma cells, examples of cancer stem cells. A comparison of these profiles was made with published data on human embryonic stem cells to determine any similarities and differences that may be of importance. The methodology involved performing expression array analysis of differentiated and undifferentiated human pluri- and nullipotent embryonal carcinoma cell lines in triplicate using Applied Biosystems technology. Data was analysed using a combination of R and Spotfire® software programs. Differentiated and undifferentiated array data was compared for each cell line to obtain a list of genes differentially expressed upon differentiation as well as a comparison of pluri- and nullipotent cell lines. Functional analysis of the data was performed using PANTHER. Validation of the data was performed by selecting a subset of genes and using TaqMan® chemistry from Applied Biosystems. The comparison with human ES data involved a literature search for expression array experiments involving human ES cells. The data produced a unique transcriptome profile of 249 differentially expressed genes upon differentiation of hEC cells at 3 days. A further profile of nullipotency was obtained by comparing nullipotent and pluripotent EC cells. The data presented has produced some novel targets that may be worth pursuing in the drive to find new more successful and effective cancer treatments.

## 6.2 Introduction

The concept of a cancer stem cell has been in existence for a long time, as far back as 1974 as proposed by Pierce (Pierce, 1974) and is gaining widespread acknowledgement as more and more published studies provide evidence of their existence in a variety of tumours including prostate, colon and breast cancer and many more (Al-Hajj et al., 2003; Collins et al., 2005;

O'Brien et al., 2007). Much of the initial work on the concept of cancer stem cells was documented in haematological malignancies such as AML and CML (Lapidot et al., 1994; Bonnet et al., 1997). The defining features of normal stem cells are their ability to self-renew and to generate differentiated progeny, properties shared by cancer stem cells.

Human embryonic stem cell (hESC) research is a vast and problematic area that could potentially reveal much information to further our understanding of the basic mechanisms of human development as well as the ultimate development of novel therapeutic agents and methods. However, there are a number of stumbling blocks including ethical and political dilemmas regarding the use of human embryonic stem cells for research purposes. Currently, much of the hESC work is done in the United States of America and in the United Kingdom. However, embryonal carcinoma cell lines were derived in both mouse and human prior to the derivation of ES cell lines and made an immense contribution to our knowledge of early cell development. Human EC cell lines have several advantages including their ability to grow without feeder cell layers, they are relatively easy to passage and they resist spontaneous differentiation.

Teratocarcinomas are malignant germ cell neoplasms that occur in the gonads and are largely composed of undifferentiated cells with pluripotent capabilities (embryonal carcinoma cell component) and differentiated cells such as squamous epithelium and neuropil (teratomatous component). The embryonal carcinoma component of these tumours is often referred to as the classical stem cell tumour as it possesses the properties of normal stem cells, namely, self-renewal and differentiation capabilities (Andrews, 2002). Within this chapter, an *in vitro* model of teratoma tumourigenesis was established using human embryonal carcinoma cell lines - Ntera-2/D1, a pluripotent embryonal carcinoma cell line derived from a testicular germ cell neoplasm by Professor Peter W. Andrews, capable of differentiation upon stimulation with retinoic acid and the second cell line used was 2102Ep, a nullipotent embryonal carcinoma cell line, which shows little response to retinoic acid and maintains an undifferentiated status also developed by Professor Peter W. Andrews (Andrews et al., 1982). Josephson et al, 2007, carried out an in depth analysis of 2102Ep and found it to be an easily maintained cell line that expressed the typical cell surface and nuclear markers of undifferentiated hESC (Josephson et al., 2007). The global gene expression profile also highly

correlated with undifferentiated hESC so his group have proposed it as a reference cell line for human ESC research.

An interesting line of research also relates to the fundamental differences between the nullipotent and pluripotent embryonal carcinoma cells and the mechanisms preventing nullipotent cells from differentiating as this may have relevance to cancer development as well as treatment. The stimulation and purposeful direction of tumours to differentiate has been used previously as a successful cancer therapy (Dean et al., 2005; Huff et al., 2006). Teratomas and the differentiated component of teratocarcinomas have been recognised to behave in a benign manner. Thus the differentiation status appears to be important in dictating the behaviour of a tumour and further important information may be gained from the comparison of nullipotent and pluripotent tumour cells.

This chapter looks at the human model of teratoma tumourigenesis in order to study human cancer stem cells and their fundamental characteristics of self-renewal and differentiation to further understand the carcinogenic process. Ultimately, the specific targeting of cancer stem cells and their complete elimination is essential to improve cancer prognosis and allow recurrence-free survival.

### 6.3 Aims

The aims of this chapter are to use the teratoma tumourigenesis cancer stem cell model to:

1. Perform a comparison of human pluripotent embryonal carcinoma cells to nullipotent embryonal carcinoma cells to elucidate the characteristics of the undifferentiated pluripotent stem cell state.
2. Establish the pathways involved and the differences between human pluripotent and nullipotent embryonal carcinoma differentiation with a focus on early differentiation.
3. Compare the data obtained with published literature on human embryonic stem cells to look at normal versus 'cancer' stem cells.
4. Compare the human model of teratoma tumourigenesis with the mouse model established in Chapters 4 and 5.

## 6.4 Materials and methods

### 6.4.1 Sample collection

The following two cell lines; Ntera-2 clone D1 – NT2/D1, human pluripotent teratocarcinoma cell line and the human nullipotent teratocarcinoma cell line, 2102Ep were cultured as per protocol (see Chapter 2, section 2.2). Cells were harvested at day 0 (1st passage) in triplicate. Each cell line was then allowed to differentiate following treatment with retinoic acid as per protocol for 3 days. Cells were harvested following 3 days differentiation in triplicate. Cells were centrifuged and all spent medium was removed. Harvested cell samples were frozen at -80°C until RNA extraction was performed.

Samples were labelled as followed:

*1. HP = Ntera-2 clone D1 – NT2/D1, human pluripotent teratocarcinoma cell line.*

a. HP1 – 3: Samples in triplicate from day 0 (1st passage).
b. HP4 – 6: Samples in triplicate from day 3, undifferentiated (2nd passage).
c. HP7 – 9: Samples in triplicate from day 3, differentiated.

*2. HC = 2102Ep, human nullipotent teratocarcinoma cell line.*

a. HC1 - 3: Samples in triplicate from day 0 (1st passage).
b. HC4 – 6: Samples in triplicate from day 3 undifferentiated (2nd passage).
c. HC7 – 9: Samples in triplicate from day 3 differentiated.

### 6.4.2 RNA extraction and in vitro transcription (IVT)

RNA extraction and in vitro transcription were performed using Qiagen kits and Applied Biosystems technology as described in Chapter 2, section 2.3 and 2.5. The samples with their associated concentrations are listed in Table 6.2. Array hybridisation and reading of arrays was performed using Applied Biosystems technology as outlined in detail in Chapter 2, section 2.5.

### 6.4.3 Data analysis

The AB1700 package for R software (a free language and environment for statistical computing and graphics, R Development Core Team, 2004) was employed to filter data using a signal/noise ratio threshold > 3 in at least one sample. It read the output from the AB1700 software with normalisation of data and performance of *t*-test and fold change with graphics to visualise *t*-test results. Fold change values were calculated for each gene filtered. Data was imported into Spotfire® (Spotfire AB, Sweden) for further analysis. A number of comparisons were performed. Differentiated and undifferentiated array data was compared for each cell line to generate a list of differentially expressed genes responsible for differentiation in both states. Undifferentiated array data from both nulli- and pluripotent cell lines was also compared to produce a list of differentially expressed genes. Only genes with a fold change ≥ 2 and a false discovery rate (FDR) of ≤ 0.05 were considered (Reiner et al., 2003), in the first instance. The gene ontology site, PANTHER, was used to identify the pathways, biological processes and molecular functions associated with significant genes. Hierarchical clustering of data was performed using Cluster 3.0 and Treeview 1.04 as described in Chapter 2, section 2.5.4.4 and 2.5.4.5.

**Table 6.1: Groups used for analysis of microarray data**

| Analysis No. | Class 1 | Class 2 |
|---|---|---|
| 1 | Undifferentiated HP cells (Day 0) | Differentiated HP cells (Day 3) |
| 2 | 'Undifferentiated' HC cells (Day 0) | 'Differentiated' HC cells (Day 3) |
| 3 | Undifferentiated HP cells | Undifferentiated HC cells |

### 6.4.4 Quantitative Real-time PCR analysis

Two μg of total RNA from each of the cell line samples was converted to cDNA using the High Capacity cDNA Archive kit (Applied Biosystems) following the manufacturer's protocol. The cDNA was used for quantitative real-time PCR amplification with TaqMan® chemistry (Applied Biosystems) using 22 pre-designed TaqMan® Gene Expression Assays from Applied Biosystems. Values were normalised relative to human GAPDH.

## 6.5 Results

### 6.5.1 RNA analysis

Quality and quantity of 18 total RNA samples was assessed by gel electrophoresis, spectrophotometry and using the Agilent 2100 Bioanalyser prior to use in expression array experiments, see Chapter 2, section 2.3.3 and 2.3.4. Concentrations of the total extracted RNA and amount of RNA used in RT-IVT reactions are shown in Table 6.2.

**Table 6.2: Concentrations (µg/µl) for samples used in Expression array analysis.**

| Sample | | Total RNA Conc. (µg/µl) | Amount for RT-IVT (µg) | Sample | | Total RNA Conc. (µg/µl) | Amount for RT-IVT (µg) |
|---|---|---|---|---|---|---|---|
| No | ID | | | No | ID | | |
| 1 | HP1 | 0.38 | 2 | 10 | HC1 | 0.60 | 2 |
| 2 | HP2 | 0.70 | 2 | 11 | HC2 | 0.60 | 2 |
| 3 | HP3 | 0.64 | 2 | 12 | HC3 | 0.60 | 2 |
| 4 | HP4 | 0.41 | 2 | 13 | HC4 | 0.90 | 2 |
| 5 | HP5 | 0.54 | 2 | 14 | HC5 | 0.70 | 2 |
| 6 | HP6 | 0.64 | 2 | 15 | HC6 | 0.50 | 2 |
| 7 | HP7 | 0.53 | 2 | 16 | HC7 | 0.80 | 2 |
| 8 | HP8 | 0.64 | 2 | 17 | HC8 | 0.60 | 2 |
| 9 | HP9 | 0.66 | 2 | 18 | HC9 | 0.60 | 2 |

### 6.5.2 cRNA analysis post RT-IVT

Post RT-IVT, cRNA was assessed qualitatively and quantitatively prior to array hybridisation. cRNA concentration was also assessed prior to array hybridisation with 10µg of cRNA used per array (see Table 6.3).

**Table 6.3: cRNA concentration of samples used for Expression array analysis.**

| Sample | cRNA conc. (µg/µl) | Amount for Array hybridisation (µg) | Sample | cRNA conc. (µg/µl) | Amount for Array hybridisation (µg) |
|---|---|---|---|---|---|
| HP1 | 0.522 | 10 | HC1 | 0.478 | 10 |
| HP2 | 0.152 | 10 | HC2 | 0.545 | 10 |
| HP3 | 0.431 | 10 | HC3 | 0.354 | 10 |
| HP4 | 1.105 | 10 | HC4 | 1.089 | 10 |
| HP5 | 0.437 | 10 | HC5 | 0.247 | 10 |
| HP6 | 0.314 | 10 | HC6 | 0.290 | 10 |
| HP7 | 0.406 | 10 | HC7 | 0.470 | 10 |
| HP8 | 0.310 | 10 | HC8 | 0.220 | 10 |
| HP9 | 0.253 | 10 | HC9 | 0.378 | 10 |

### 6.5.3 Statistical analysis

#### *6.5.3.1 Normalisation*

A 5% trimmed mean was the main normalisation method used in this study; this method produces reliable results when analysing low signal intensity genes. Figure 6.1 shows MA plots after normalisation for the biological replicates HP1 – HP6, HP7 – HP9, HC1 – HC6 and HC7 – HC9. Scatter plots for replicates are symmetrical around zero and show no appreciable difference. The three replicates from each passage show closer similarities than a comparison of a replicate from each of the two passages.

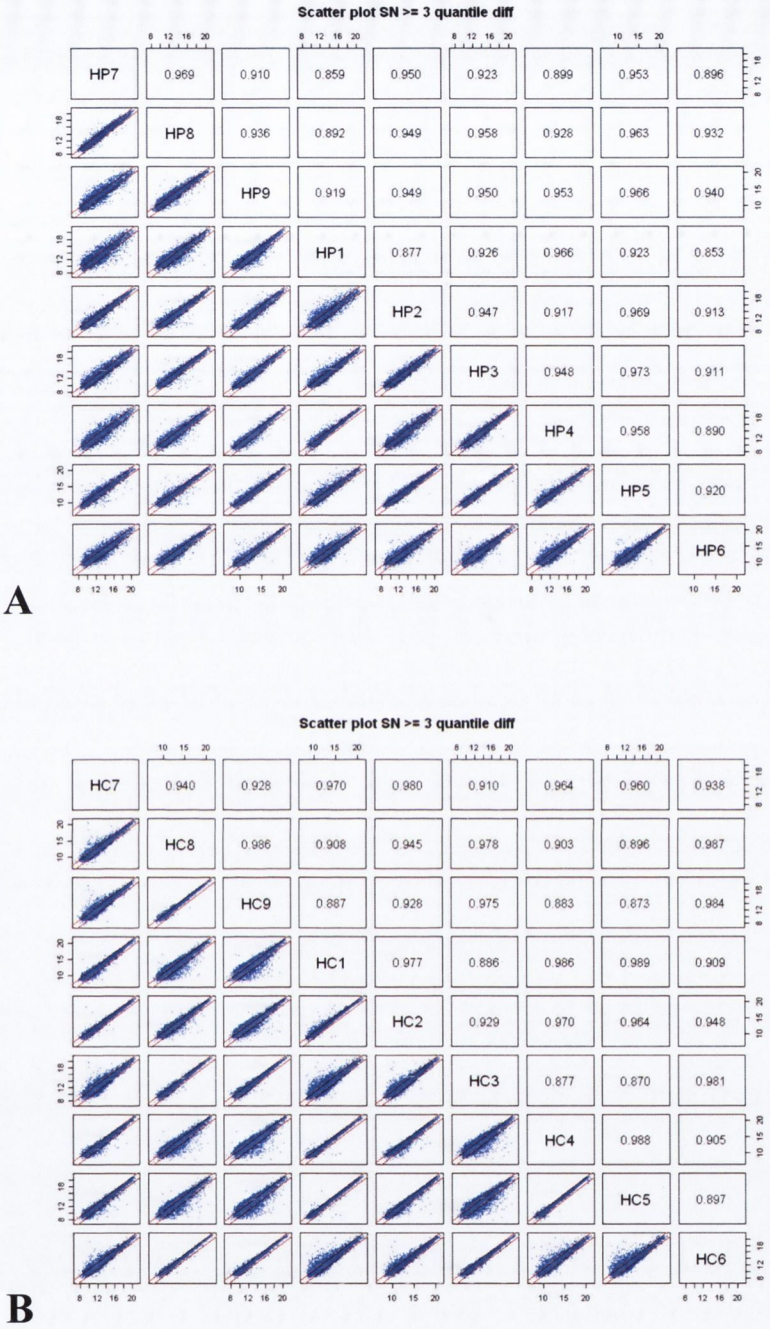



**Figure 6.1: Scatter plots for A) HP and B) HC cell lines.**

### *6.5.3.2 Filtering genes*

Step-wise filtering of gene probes was performed for all analyses as follows:

1) During normalisation in ab1700gui, probes with a signal-to-noise ratio of less than 3 (S/N<3) in all arrays were eliminated.

2) Detectable genes were then analysed using a t test in ab1700gui and a correction for multiple testing. Genes with an FDR ≤ 0.05 were considered to show significant differential expression between two classes of analysis. Genes that were not significantly differentially expressed after t test or adjustment for multiple comparisons were eliminated.
3) Genes less than two-fold up- or down-regulated (FC < 2) between the two classes were eliminated.

### 6.5.4 Analysis of Data Groups

#### *6.5.4.1 Analysis 1: Comparison of undifferentiated HP cells (day 0) v differentiated HP cells (day 3)*

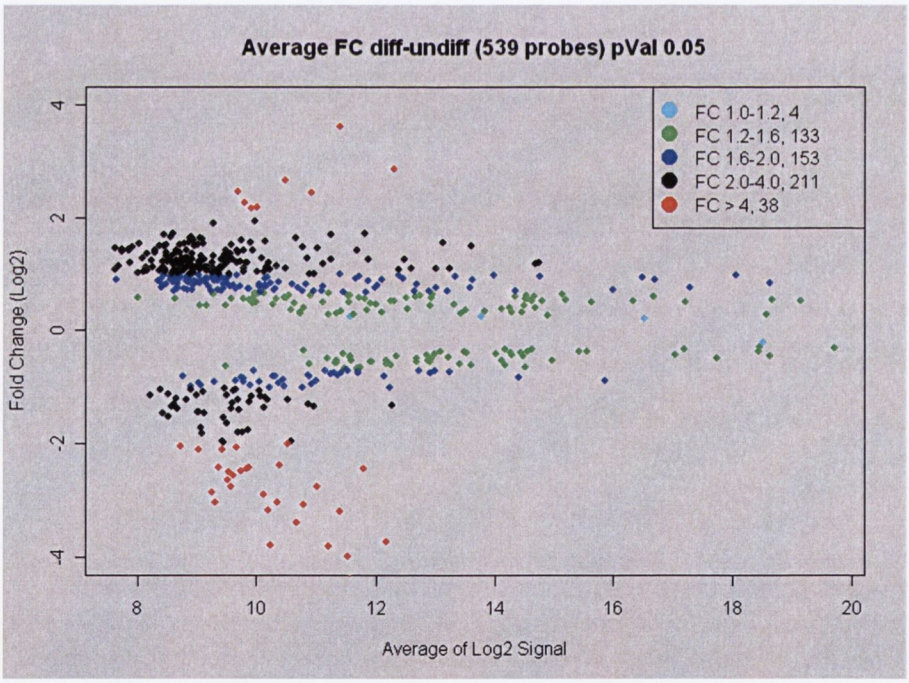



**Figure 6.2: Differentially expressed probes with a p value of < 0.05 in a comparison of differentiated and undifferentiated HP cells.**

As there were no genes obtained with a FDR ≤ 0.05, 0.25 or 0.5, it was decided to be less stringent in the analysis, and therefore p values were considered with only those genes with a p value < 0.05 considered for further analysis.

Differentially expressed genes with a p value < 0.05 539 genes

- 329 (61.0%) of this list of genes is known and named.
- The remainder, 210 (38.9%) are unknown.

Of these: 367 are upregulated upon differentiation - 68.1%

172 are downregulated upon differentiation - 31.9%

Differentially expressed genes with a p value < 0.05, fold change > 2 and < 0.5:

249 genes.

- 144 (57.8%) of this list of genes is known and named.
- 105 (42.2%) of this list of genes is unknown.

Of these: 173 are upregulated upon differentiation - 69.5%

76 are downregulated upon differentiation - 30.5%

- 36 (47.4%) of these genes are unknown.

**Table 6.4: Top 10 genes upregulated upon differentiation of HP cell line.**

| Top 10 Upregulated genes in differentiated HP cells | | | | | | |
|---|---|---|---|---|---|---|
| **FC** | **Log2(FC)** | **p value** | **FDR** | **FC Bin** | **Gene_Symbol** | **Gene_Name** |
| 12.303 | 3.621 | 0.034 | 0.939 | FC > 4 | *PNCK* | pregnancy upregulated non-ubiquitously expressed CaM kinase |
| 7.300 | 2.868 | 0.045 | 0.939 | FC > 4 | *POLI* | polymerase (DNA directed) iota |
| 5.455 | 2.447 | 0.033 | 0.939 | FC > 4 | *TSPAN18* | tetraspanin 18 |
| 4.830 | 2.272 | 0.027 | 0.939 | FC > 4 | *HEY2* | hairy/enhancer-of-split related with YRPW motif 2 |
| 4.613 | 2.206 | 0.013 | 0.939 | FC > 4 | *MAPK13* | mitogen-activated protein kinase 13 |
| 4.543 | 2.184 | 0.023 | 0.939 | FC > 4 | *CALB2* | calbindin 2, 29kDa (calretinin) |
| 3.834 | 1.939 | 0.011 | 0.939 | FC 2.0-4.0 | *POU3F1* | POU domain, class 3, transcription factor 1 |
| 3.287 | 1.717 | 0.006 | 0.939 | FC 2.0-4.0 | *C19orf28* | chromosome 19 open reading frame 28 |
| 3.286 | 1.716 | 0.048 | 0.939 | FC 2.0-4.0 | *PLAGL1* | pleiomorphic adenoma gene-like 1 |
| 3.263 | 1.706 | 0.002 | 0.939 | FC 2.0-4.0 | *AMN* | amnionless homolog (mouse) |

**Table 6.5: Top 10 genes downregulated upon differentiation of HP cell line.**

| Top 10 Downregulated in differentiated HP cells | | | | | | |
|---|---|---|---|---|---|---|
| FC | Log2(FC) | p value | FDR | FC Bin | Gene_Symbol | Gene_Name |
| 0.035 | -4.823 | 0.009 | 0.939 | FC > 4 | *TMC2* | transmembrane channel-like 2 |
| 0.071 | -3.815 | 0.009 | 0.939 | FC > 4 | *IL17RB* | interleukin 17 receptor B |
| 0.073 | -3.780 | 0.008 | 0.939 | FC > 4 | *Cep192* | centrosomal protein 192 kDa |
| 0.075 | -3.736 | 0.020 | 0.939 | FC > 4 | *GPR103* | G protein-coupled receptor 103 |
| 0.109 | -3.201 | 0.024 | 0.939 | FC > 4 | *ZNF598* | zinc finger protein 598 |
| 0.110 | -3.181 | 0.021 | 0.939 | FC > 4 | *C9orf52* | chromosome 9 open reading frame 52 |
| 0.119 | -3.069 | 0.036 | 0.939 | FC > 4 | *na* | hypothetical LOC389025 |
| 0.149 | -2.749 | 0.011 | 0.939 | FC > 4 | *KIF2B* | kinesin family member 2B |
| 0.161 | -2.632 | 0.016 | 0.939 | FC > 4 | *FLJ90650* | laeverin |
| 0.181 | -2.469 | 0.012 | 0.939 | FC > 4 | *HCN4* | hyperpolarization activated cyclic nucleotide-gated potassium channel 4 |

**PANTHER analysis:**

None of the pathways, biological processes and molecular functions was significantly over represented in the differentially expressed genes when compared with the AB1700 reference list of human genes. This analysis was based on the list of 539 genes so includes genes with fold changes < 2. The pathways, biological processes and molecular functions of this list were examined.

**Table 6.6: Differentially expressed genes in Ntera2 cells and their associated pathways taken from PANTHER. The majority of the list of 539 differentially expressed genes is unclassified (84%). A selection of the pathways is highlighted in this table.**

| Pathway | Number of genes | Gene Symbols |
|---|---|---|
| **Unclassified** | 453 | |
| **Angiogenesis** | 10 | PAK2, SPHK2, FGFR3, PDLIM7, DVL2, JAK1, PRKCQ, MAPK13, 3 unknown |
| **JAK/STAT** | 2 | JAK1, MAPK13 |
| **Wnt signalling** | 5 | DVL2, PRKCQ, SMO, HELLS, unknown |
| **Hedgehog signalling** | 3 | GLI1, PRKAR2A, SMO |
| **TGF-β signalling** | 5 | FOXK2, FOXP4, FOX03A, MAPK13, unknown |
| **Notch signalling** | 3 | HEY2, CIR, LNX2 |

Outside of the unclassified group accounting for 84% of the gene list, angiogenesis featured highest with 10 genes in this list, 3 of which are unknown.

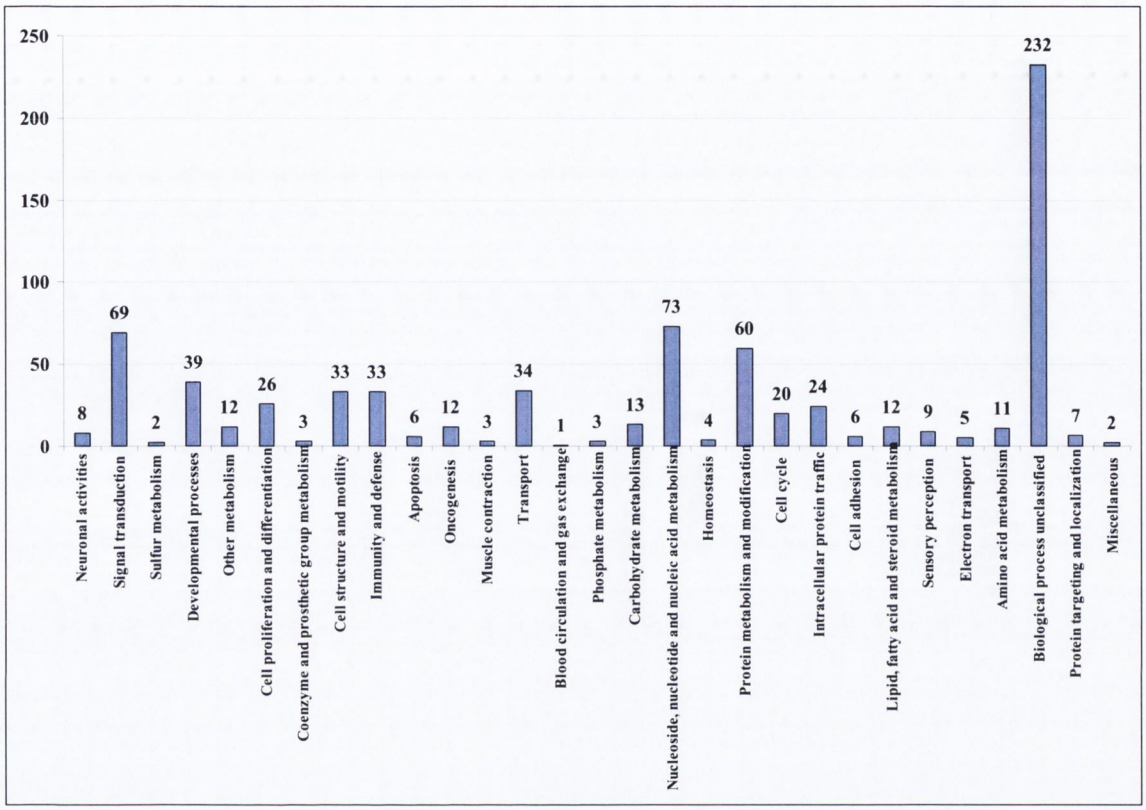



**Figure 6.3: Biological processes involved in the differentiation of Ntera2 (HP) cells. Many are unclassified (232, 43%), protein metabolism, nucleoside, nucleotide and nucleic acid metabolism and signal transduction contain high numbers of differentially expressed genes.**

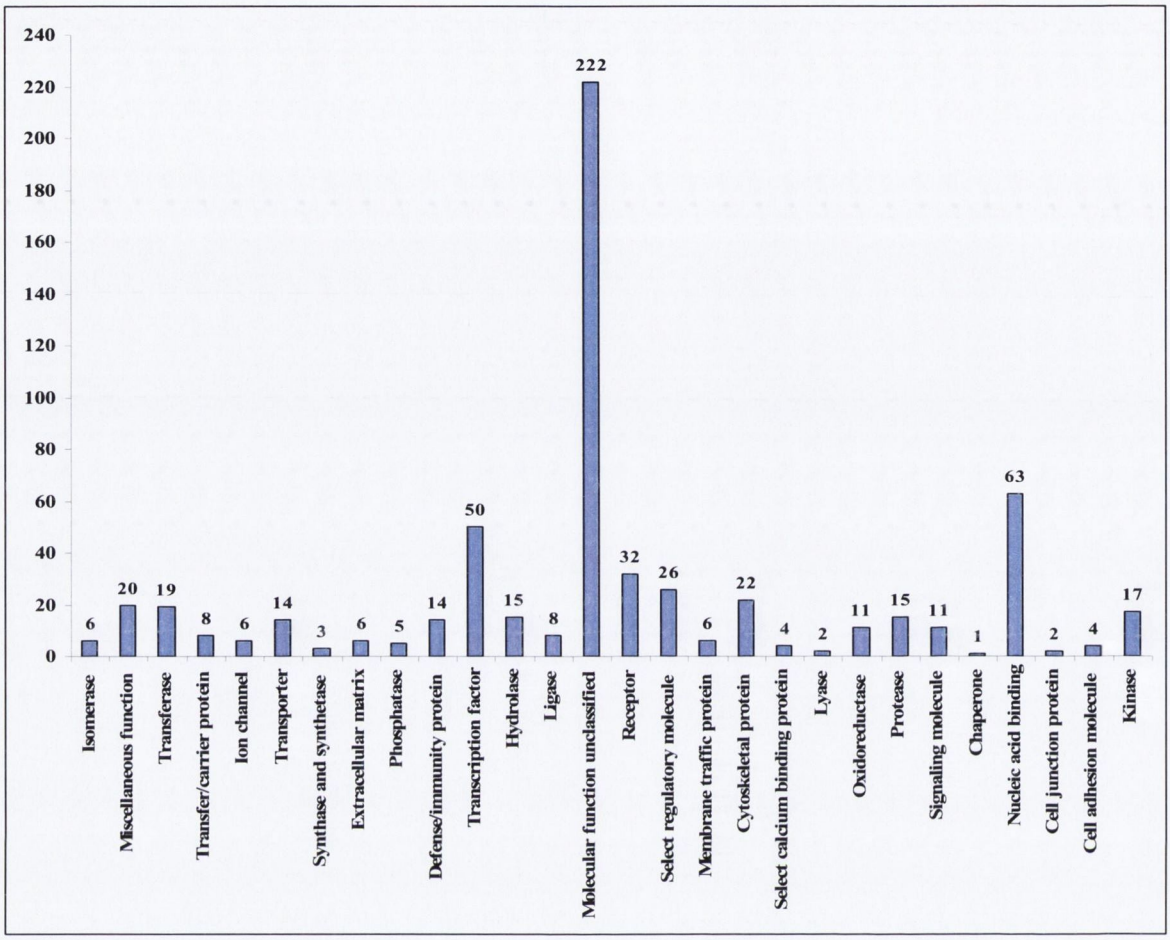



**Figure 6.4: Molecular functions: Of the list of 539 genes differentially expressed upon differentiation of HP cell, many are in the unclassified category (222, 41.2%) with nucleic acid binding, and transcription factors, receptors and select regulatory molecules next in line.**

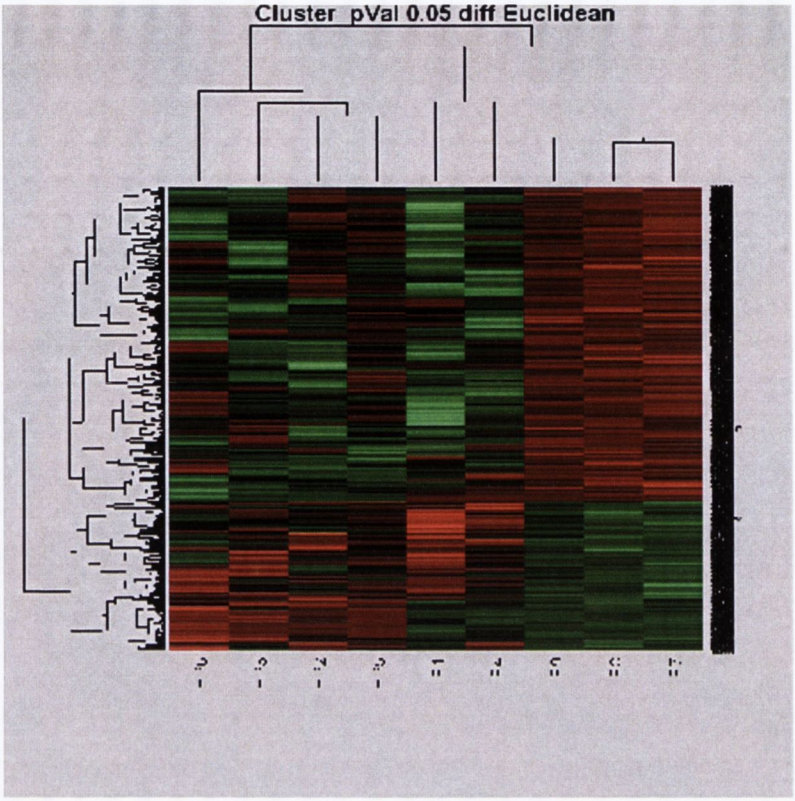



**Figure 6.5: Hierarchical clustering of HP differentiated cells versus undifferentiated cells.**

As expected, the arrays labelled HP7-9 (differentiated samples) have clustered together with HP1-HP6 (undifferentiated) samples also clustering together as expected. Interestingly, arrays HP1 and 4 were more similar to the differentiated samples than their remaining undifferentiated counterparts.

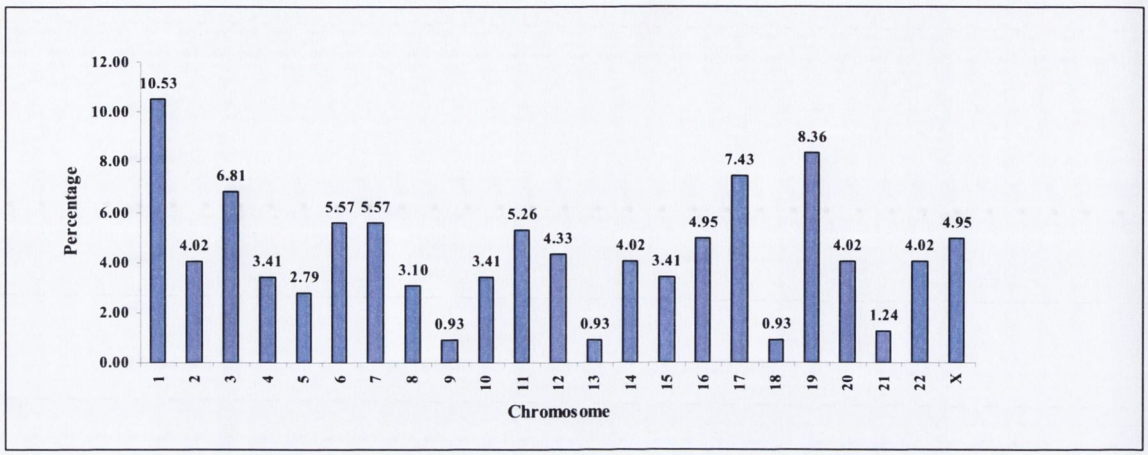



**Figure 6.6: Percentage of genes per chromosome (x axis = chromosome number, y axis = percentage) of the 539 differentially expressed genes upon differentiation of HP cells. Chromosome 1 has the highest number of genes involved at 10.53%.**

**Table 6.7: Signal Transduction – 32 genes differentially expressed upon differentiation of HP cells at 3 days, specifically involved in signal transduction (when only those with a fold change > 2 or < 0.5 are considered).**

| Genes differentially expressed in HP cells upon differentiation involved in: Signal transduction | | | | | |
|---|---|---|---|---|---|
| FC (diff/undiff) | Log2(FC) | p value | FDR (BH) | FC Bin | Gene Symbol |
| 12.303 | 3.621 | 0.034 | 0.939 | FC > 4 | PNCK |
| 6.367 | 2.671 | 0.038 | 0.939 | FC > 4 | Unknown |
| 4.613 | 2.206 | 0.013 | 0.939 | FC > 4 | MAPK13 |
| 4.543 | 2.184 | 0.023 | 0.939 | FC > 4 | CALB2 |
| 3.263 | 1.706 | 0.002 | 0.939 | FC 2.0-4.0 | AMN |
| 2.759 | 1.464 | 0.047 | 0.939 | FC 2.0-4.0 | PVRL2 |
| 2.698 | 1.432 | 0.036 | 0.939 | FC 2.0-4.0 | FOXP4 |
| 2.669 | 1.416 | 0.013 | 0.939 | FC 2.0-4.0 | GPR30 |
| 2.530 | 1.339 | 0.014 | 0.939 | FC 2.0-4.0 | Unknown |
| 2.416 | 1.273 | 0.035 | 0.939 | FC 2.0-4.0 | AKT1S1 |
| 2.408 | 1.268 | 0.046 | 0.939 | FC 2.0-4.0 | JUP |
| 2.346 | 1.230 | 0.013 | 0.939 | FC 2.0-4.0 | WARP |
| 2.314 | 1.210 | 0.018 | 0.939 | FC 2.0-4.0 | DVL2 |
| 2.297 | 1.200 | 0.006 | 0.939 | FC 2.0-4.0 | SSB3 |
| 2.228 | 1.156 | 0.021 | 0.939 | FC 2.0-4.0 | SEMA7A |
| 2.191 | 1.131 | 0.039 | 0.939 | FC 2.0-4.0 | Unknown |
| 2.173 | 1.120 | 0.042 | 0.939 | FC 2.0-4.0 | SLB |
| 2.149 | 1.104 | 0.006 | 0.939 | FC 2.0-4.0 | CAMK1G |
| 2.111 | 1.078 | 0.010 | 0.939 | FC 2.0-4.0 | PPGB |
| 2.049 | 1.035 | 0.035 | 0.939 | FC 2.0-4.0 | PRKAR2A |
| 0.476 | -1.072 | 0.008 | 0.939 | FC 2.0-4.0 | SPPL2A |
| 0.456 | -1.132 | 0.030 | 0.939 | FC 2.0-4.0 | TEC |
| 0.450 | -1.152 | 0.020 | 0.939 | FC 2.0-4.0 | GRM5 |
| 0.447 | -1.161 | 0.014 | 0.939 | FC 2.0-4.0 | PRKCQ |
| 0.434 | -1.205 | 0.041 | 0.939 | FC 2.0-4.0 | |
| 0.410 | -1.285 | 0.006 | 0.939 | FC 2.0-4.0 | OLFML2A |
| 0.319 | -1.646 | 0.039 | 0.939 | FC 2.0-4.0 | MME |
| 0.296 | -1.757 | 0.005 | 0.939 | FC 2.0-4.0 | IL2RG |
| 0.282 | -1.827 | 0.017 | 0.939 | FC 2.0-4.0 | Unknown |
| 0.282 | -1.827 | 0.017 | 0.939 | FC 2.0-4.0 | Unknown |
| 0.258 | -1.956 | 0.030 | 0.939 | FC 2.0-4.0 | P2RY1 |
| 0.233 | -2.099 | 0.031 | 0.939 | FC > 4 | AKAP5 |
| 0.191 | -2.386 | 0.042 | 0.939 | FC > 4 | OPTC |

### *6.5.4.2 Analysis 2: Comparison of undifferentiated HC cells (day 0) v differentiated HC cells (day 3)*

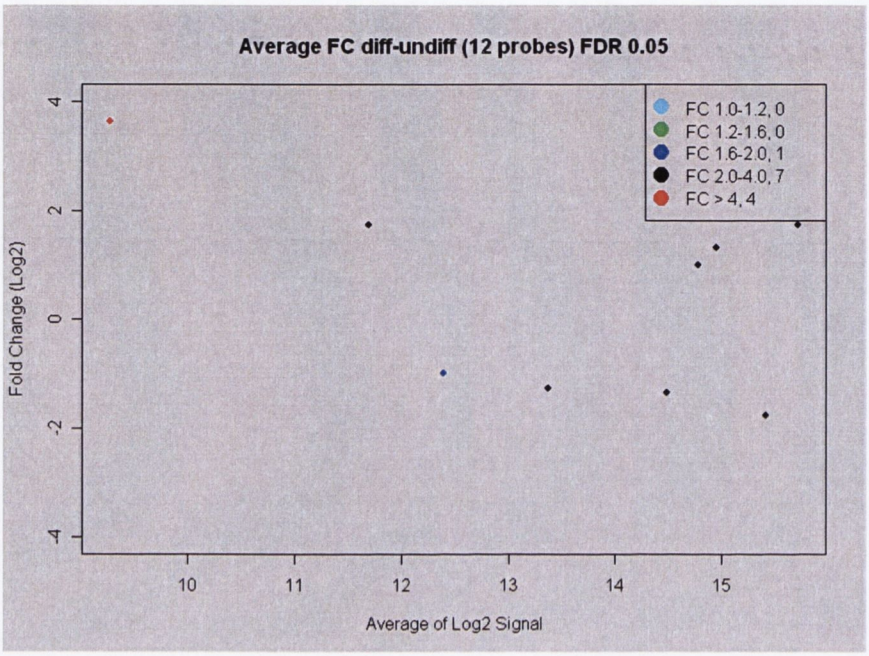



**Figure 6.7: Number of gene probes differentially expressed between 'undifferentiated' and 'differentiated' HC cells with a FDR ≤ 0.05.**

**Table 6.8: The 12 differentially expressed probes with FDR ≤ 0.05 upon 'differentiation' of HC (nulli) cells.**

| Genes differentially expressed in HC (Nullipotent) cell differentiation | | | | | |
|---|---|---|---|---|---|
| | FC (diff/undiff) | p value | FDR | Gene Symbol | Gene Name |
| **1** | 192.638 | 2.66E-08 | 0.00048 | **CYP26A1** | cytochrome P450, family 26, subfamily A, polypeptide 1 |
| **2** | 50.908 | 4.51E-07 | 0.004063 | **ZNF503** | zinc finger protein 503 |
| **3** | 0.502 | 8.52E-07 | 0.005112 | **LHX6** | LIM homeobox 6 |
| **4** | 21.989 | 1.48E-06 | 0.006649 | **HOXB5** | homeo box B5 |
| **5** | 12.636 | 4.38E-06 | 0.015778 | **CD22** | CD22 antigen |
| **6** | 2.516 | 7.01E-06 | 0.021028 | | |
| **7** | 3.346 | 1.06E-05 | 0.027323 | **NMU** | neuromedin U |
| **8** | 3.342 | 2.03E-05 | 0.039054 | | |
| **9** | 2.019 | 1.79E-05 | 0.039054 | **FLJ11286** | hypothetical protein FLJ11286 |
| **10** | 0.414 | 2.36E-05 | 0.039054 | **GPR143** | G protein-coupled receptor 143 |
| **11** | 0.396 | 2.48E-05 | 0.039054 | **FOXD1** | forkhead box D1 |
| **12** | 0.294 | 2.60E-05 | 0.039054 | **na** | hypothetical LOC388638 |

Four genes are downregulated in 'differentiated' HC cells while the remaining 8 are upregulated in 'differentiated' HC cells.

### *6.5.4.3 Analysis 3: Comparison of undifferentiated HC (nulli) cells (day 0) v undifferentiated HP (pluri) cells (day 0)*

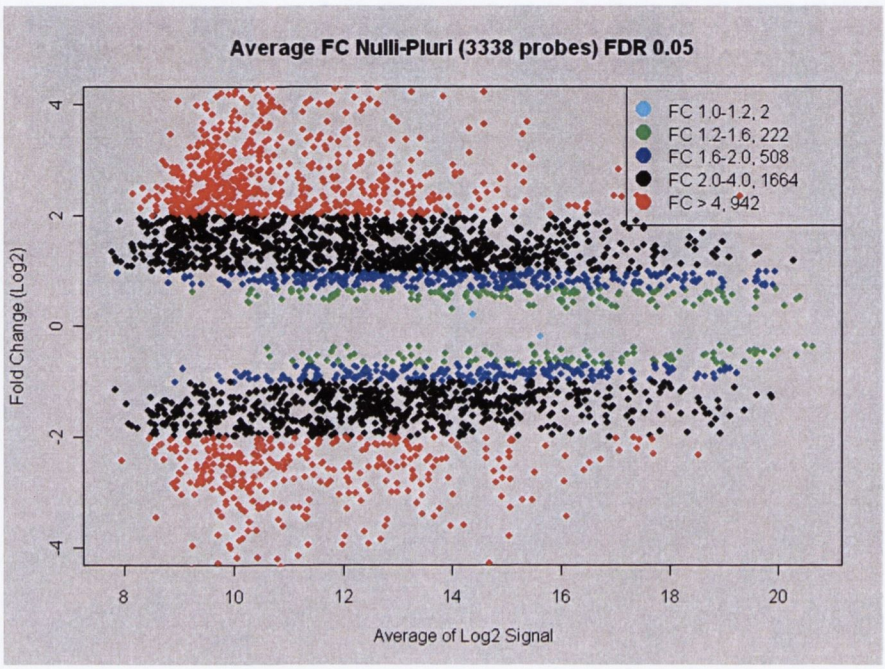



**Figure 6.8: Number of gene probes differentially expressed between undifferentiated HC (nulli) and undifferentiated HP (pluri) cells with a FDR ≤ 0.05.**

Genes differentially expressed in nullipotent compared with pluripotent cells:

3,338 genes with a FDR ≤ 0.05

- 2,607 genes with a FDR ≤ 0.05 and fold change of ≥ 2 or < 0.5.
- Of these:
    - 1,031 genes (39.6%) are downregulated.
    - 1,575 genes (60.4%) are upregulated.

**Table 6.9: Top ten upregulated in Nullipotent cells compared with pluripotent cells.**

| Top 10 Genes upregulated in HC (Nullipotent) v HP (pluripotent) cells | | | | | | |
|---|---|---|---|---|---|---|
| FC (Nulli/Pluri) | Log2(FC) | p value | FDR (BH) | FC Bin | Gene Symbol | Gene Name |
| 1915.01 | 10.90 | 7.28E-11 | 1.33E-06 | FC > 4 | na | similar to developmental pluripotency associated 5; embryonal stem cell specific gene 1 |
| 202.35 | 7.66 | 1.35E-06 | 1.68E-04 | FC > 4 | LCP1 | lymphocyte cytosolic protein 1 (L-plastin) |
| 140.10 | 7.13 | 6.89E-07 | 1.11E-04 | FC > 4 | FLJ32942 | hypothetical protein FLJ32942 |
| 83.66 | 6.39 | 1.19E-08 | 9.10E-06 | FC > 4 | HORMAD1 | HORMA domain containing 1 |
| 81.63 | 6.35 | 4.10E-08 | 2.21E-05 | FC > 4 | CDH1 | cadherin 1, type 1, E-cadherin (epithelial) |
| 77.31 | 6.27 | 3.53E-09 | 4.04E-06 | FC > 4 | | |
| 68.35 | 6.09 | 2.01E-05 | 9.26E-04 | FC > 4 | CRI1 | CREBBP/EP300 inhibitor 1 |
| 61.58 | 5.94 | 3.24E-08 | 1.98E-05 | FC > 4 | C14orf29 | chromosome 14 open reading frame 29 |
| 61.10 | 5.93 | 2.40E-05 | 1.04E-03 | FC > 4 | THUMPD2 | THUMP domain containing 2 |
| 58.69 | 5.87 | 4.92E-05 | 1.61E-03 | FC > 4 | | |

**Table 6.10: Top ten downregulated in Nullipotent cells compared with pluripotent cells.**

| Top 10 Downregulated genes in HC (Nullipotent) v HP (Pluripotent) cells | | | | | | |
|---|---|---|---|---|---|---|
| FC (Nulli/Pluri) | Log2(FC) | p value | FDR (BH) | FC Bin | Gene_Symbol | Gene_Name |
| 0.0088 | -6.8308 | 9.00E-09 | 7.84E-06 | FC > 4 | | |
| 0.0126 | -6.3083 | 3.07E-08 | 1.94E-05 | FC > 4 | MYH6 | myosin, heavy polypeptide 6, cardiac muscle, alpha (cardiomyopathy, hypertrophic 1) |
| 0.0205 | -5.6078 | 6.23E-07 | 0.000106 | FC > 4 | LHX2 | LIM homeobox 2 |
| 0.0243 | -5.3658 | 2.87E-07 | 6.48E-05 | FC > 4 | ECAT8 | ES cell associated transcript 8 |
| 0.0263 | -5.2494 | 6.94E-07 | 0.000111 | FC > 4 | FEZ1 | fasciculation and elongation protein zeta 1 (zygin I) |
| 0.0277 | -5.1727 | 5.70E-07 | 0.000101 | FC > 4 | | |
| 0.0294 | -5.0858 | 0.00011 | 0.00279 | FC > 4 | | |
| 0.0303 | -5.0422 | 1.08E-07 | 3.95E-05 | FC > 4 | PIWIL2 | piwi-like 2 (Drosophila) |
| 0.0329 | -4.9272 | 1.35E-09 | 2.74E-06 | FC > 4 | HMGA2 | high mobility group AT-hook 2 |
| 0.0340 | -4.8789 | 7.37E-10 | 1.73E-06 | FC > 4 | RARRES2 | retinoic acid receptor responder (tazarotene induced) 2 |

## PANTHER analysis

Following importing the probe IDs into PANTHER of the differentially expressed gene list, a comparison was made with a reference list of human genes to determine which pathways, biological processes and molecular functions were over represented in this list compared with a reference list.

| ***Pathway analysis:*** | ***P value*** | ***No. of genes*** |
|---|---|---|
| Cadherin signalling pathway | 6.46E-06 | 43 |
| Wnt signalling pathway | 5.74E-03 | 61 |
| Pyridoxal phosphate salvage pathway | 8.91E-03 | 4 |
| Vitamin B6 metabolism | 2.67E-02 | 5 |

| ***Biological process:*** | ***P value*** | ***No. of genes*** |
|---|---|---|
| Homeostasis | 4.71E-04 | 41 |
| Other homeostasis activities | 6.93E-04 | 23 |
| Protein modification | 4.72E-03 | 151 |
| Transport | 1.08E-02 | 165 |
| Lipid and fatty acid binding | 2.58E-02 | 10 |

| ***Molecular function:*** | ***P value*** | ***No. of genes*** |
|---|---|---|
| Storage function | 4.11E-09 | 15 |
| Select regulatory molecule | 8.72E-04 | 154 |
| Ligase | 1.36E-03 | 69 |
| Chromatin/Chromatin binding-protein | 4.12E-02 | 31 |

*6.5.4.4 Comparison of Ntera2 and hESC data*

Another aim of this chapter was to compare the Ntera2 data obtained to hESC data in the literature. A number of papers in the literature have compared undifferentiated hESC cells to differentiated hESCs using a number of time points (usually later than 3 days as used in this study) and using a variety of gene expression analysis methods such as massively parallel signature sequencing (MPSS), serial analysis of gene expression (SAGE), expressed sequence tag (EST), large scale microarrays, focused cDNA microarrays and immunohistochemistry.

Liu et al, 2006 performed a comparison of seven hESC lines using Ilumina® bead arrays with 24,131 transcript probes (Liu et al., 2006). Further comparisons were performed between hESCs, EBs (at 14 days), human fibroblast feeder cells and undifferentiated Ntera2 cells. They found that hESCs and EBs could be distinguished from each other and produced a list of a

subset of gene probes, 44 in total, that were significantly over expressed in EBs compared with hESCs.

Comparison of this list of 44 genes with the genes found in my study to be differentially regulated in Ntera2 differentiation (differentiation at 3 days) – a comparison not performed by the authors:

- 22 (50%) were not found to be differentially regulated in Ntera2 differentiation.
- 7 (15.9%) were found to be > 2 fold over expressed upon differentiation of Ntera2 cells (although all p values were > 0.05) – See Table 6.11.
- The remaining 15 genes all had fold change values < 2 or > 0.5. All p values were also > 0.05.

**Table 6.11: List of 7 genes from Liu et al's list of 44 genes which were differentially expressed in Ntera2 differentiation. Note that all 7 genes are upregulated in differentiated Ntera2 cells.**

| Gene symbol | Gene name | Fold change | P value |
|---|---|---|---|
| **PAX6** | paired box gene 6 (aniridia, keratitis) | 5.503 | 0.288 |
| **IGFBP3** | insulin-like growth factor binding protein 3 | 5.281 | 0.178 |
| **PITX2** | paired-like homeodomain transcription factor 2 | 4.673 | 0.347 |
| **COL2A1** | collagen, type II, alpha 1 (primary osteoarthritis, spondyloepiphyseal dysplasia, congenital) | 3.200 | 0.075 |
| **BMP4** | Bone morphogenetic protein 4 | 2.829 | 0.101 |
| **APOB** | Apolipoprotein B binding protein | 2.777 | 0.340 |
| **FRZB** | Frizzled related protein | 2.002 | 0.109 |

The authors also compared undifferentiated hESCs to undifferentiated Ntera2 cells and found that there were many similarities. There were some differences which seemed to reflect the origin of the tumour cells from which the Ntera2 was derived. For example, germ cell markers

such as GAGE2, GAGE7 and GAGE8 were highly expressed in Ntera2 but absent in any of the hESC lines examined. Interestingly, they also found a significant difference in the expression of genes in the TGF-β pathway in Ntera2 such as CDKN1A, IGFBP7, NODAL and BMP2.

Brandenberger et al performed a comparison of undifferentiated and differentiated hESCs using embryoid bodies at 8 days (Brandenberger et al., 2004b). They produced a list of 672 differentially expressed genes, with 38.6% of these genes unknown. They revealed the presence of several key signalling pathways in hES cells that included FGF, WNT, NODAL and LIF pathways.

**Table 6.12: Comparison of data on differentiated versus undifferentiated hES cells from Brandenberger et al and differentiated versus undifferentiated Ntera2 cells in this thesis (Brandenberger et al., 2004b).**

| *Gene name* | *Brandenberger et al* | *My study* |
|---|---|---|
| | **LIF Pathway** | |
| **POU5F1** | Downregulated | Downregulated (1.27 fold) |
| **IL6ST** | Upregulated (112 fold at day 9) | Upregulated (1.85 fold) |
| | **WNT pathway** | |
| **WNT5A** | Upregulated | Unchanged |
| **SFRP2** | Downregulated | Downregulated (2.17 fold) |
| **CTNNB** | Upregulated | Upregulated (1.78 fold) |
| | **NODAL pathway** | |
| **TDGF1** | Downregulated | Downregulated (5.23 fold) |
| **LEFTB** | Downregulated | Unchanged |
| **FOXH1** | Downregulated | Downregulated (1.10 fold) |

Many similarities are noted in the genes described as significant by Brandenberger et al with the main differences being the level of fold change occurring which is always lower in my study, however this may be explained by the authors use of a later differentiation time point.

**Table 6.13: Genes characteristic of the undifferentiated hES cell (Brandenberger et al., 2004a) with their associated fold change values from this chapter's array experiments. * = p value < 0.05 or FDR > 0.05.**

| | Gene Symbol | Ntera2 diff v undiff | 2102Ep v Ntera2 |
|---|---|---|---|
| 1 | DNMT3B | 0.841* | 1.294* |
| 2 | POU5F1 | 1.076* | 0.824* |
| 3 | ZFP42 | Unchanged | 22.931 |
| 4 | SOX2 | 1.314* | 0.471* |
| 5 | TERF1 | 1.275* | 0.852* |
| 6 | GJA1 | 1.045* | 1.681* |
| 7 | NOL7 | 0.856* | 1.816 |
| 8 | GAL | 0.432* | 1.841** |
| 9 | UTF1 | 0.997* | 1.412* |
| 10 | LEFTB | 12.237* | 7.602* |
| 11 | LOC388638 | 0.074* | 0.706* |
| 12 | TERF2IP | 0.797* | 0.987* |
| 13 | PODXL | 2.249* | 0.404 |
| 14 | TDGF1 | 0.191* | 16.335* |
| 15 | EBAF | 17.817* | 1.972* |
| 16 | TERF2 | 1.177* | 1.157* |
| 17 | NANOG | 0.604* | 5.924* |
| 18 | NODAL | Unchanged | 17.969 |

These are 18 genes characteristic of the undifferentiated hES cell as described by Branderberger et al, 2004. Looking at the data from Ntera2 differentiation, none of the genes mentioned has a significant p value, however, 2 genes were unchanged after 3 days differentiation. Of the remainder, 8 were downregulated upon differentiation and 8 were upregulated upon differentiation. The four highlighted genes were significantly differentially expressed in 2102Ep cells (nullipotent) compared with Ntera2 cells (pluripotent) suggesting that these genes have a significant role in the maintenance of the self-renewing state.

#### *6.5.4.5 Comparison of mouse and human model*

*hES versus mES*

Bhattacharya et al 2004, presented a molecular signature of hES cells of 92 genes which were over expressed by 3 fold and enriched within hES cells (Bhattacharya et al., 2004). They compared this 92 gene list to published lists of genes over expressed in mES cells (Ivanova et al., 2002; Ramalho-Santos et al., 2002; Tanaka et al., 2002) and found a limited overlap of between 12 and 33 genes. Explanations for this low degree of concordance were their use of 3 fold over expression rather than 1.4 fold used by the other studies as well as the use of LIF

rather than fibroblast feeder layers in some mES experiments which may account for some differences.

A comparison of this 92 gene to the list of 443 genes over expressed in undifferentiated mES cells (chapter 5, section 5.5.4.3) shows only two similarities. This poor concordance may reflect the abundance of unknown genes retrieved from the mES data (133/443 or 30.02% of genes unknown or unnamed) as well as the earlier time point of differentiation of 3 days used. Their experiment used embryoid bodies over 8 days after differentiation. The authors conclude that the limited overlap between studies may suggest that truly universal stem cell markers are uncommon and may be a subset of less than 100 genes that may require a more direct comparison of purified homogenous populations of stem cells.

*Human EC v mouse EC and ES cells*

Human orthologs are given for all genes in the Applied Biosystem array output of the mouse data and likewise in the human data, mouse orthologs were given.

To compare the human and mouse EC data, two searches were done.

a. Using the mouse data to search for human orthologs.
b. Using the human data to search for mouse orthologs.

***Comparison A:***

In the comparison of differentiated SCC-PSA1 (pluripotent embryonal carcinoma) versus the undifferentiated state:

- 16,285 genes differentially expressed in total.
- 4,493 (27.6%) had no human ortholog.
- Of the remaining 11,792 genes, confining the analysis to those genes with a FDR < 0.05 and fold change < 2 or < 0.05 results in 162 genes with a human ortholog.
    - 124 (76.5%) are downregulated upon differentiation.
    - 38 (23.5%) are upregulated upon differentiation.

The human ortholog of each of these 162 genes was searched for in the differentiated Ntera2 versus undifferentiated comparison to look for any similarities or differences.

- 61 (37.7%) genes were not found to be differentially expressed in the Ntera2 comparison.
- 10 (6.2%) genes were unnamed.
- Of the remaining 91 genes, only 2 had a p value $< 0.05$.
    i. SEMA7A
        1. Ntera2: fold change 2.228.
        2. SCC-PSA1: fold change 0.275.
    ii. LOXL1
        1. Ntera2: fold change 1.619.
        2. SCC-PSA1: fold change 0.320.

*Comparison B:*

Using the 76 gene signature of undifferentiated Ntera2 cells only, these 76 genes were searched to see how many mouse orthologs.

- 38 had no mouse ortholog.
- The remaining 38 had a mouse ortholog.
    i. In comparison with SCC-PSA1 (mEC pluripotent cells), 22 (57.9%) were differentially expressed. However, none had a FDR $< 0.05$. Using a FDR $< 0.1$, results in two genes being significant, namely Fos and Egr1.

**Table 6.14: 22 mouse genes from SCC-PSA1 differentiation.**

| *Mouse Orthologs* | *ProbeID* | *FC (diff/ undiff)* | *p value* | *FDR* | *Gene_Symbol* | *Gene_Name* |
|---|---|---|---|---|---|---|
| mCG5518.1 | 463565 | 0.316 | 0.002 | 0.053 | **Fos** | FBJ osteosarcoma oncogene |
| mCG123917.1 | 524988 | 0.299 | 0.004 | 0.073 | **Egr1** | early growth response 1 |
| mCG4792.2 | 885628 | 1.385 | 0.013 | 0.127 | **Ripk4** | receptor-interacting serine-threonine kinase 4 |
| mCG19628.2 | 848724 | 0.818 | 0.016 | 0.138 | **Prkar1a** | protein kinase, cAMP dependent regulatory, type I, alpha |
| mCG16156.2 | 880942 | 1.953 | 0.043 | 0.227 | **Trf** | transferrin |
| mCG1039741.1 | 875245 | 1.789 | 0.061 | 0.267 | **Hbq1** | hemoglobin, theta 1 |
| mCG142428.1 | 863193 | 0.355 | 0.063 | 0.273 | **Gsto2** | glutathione S-transferase omega 2 |
| mCG8982.1 | 450459 | 1.368 | 0.071 | 0.286 | **1810054D07Rik** | RIKEN cDNA 1810054D07 gene |
| mCG1037923.1 | 897869 | 0.714 | 0.147 | 0.405 | **Tdgf1** | teratocarcinoma-derived growth factor |
| mCG140437 | 545994 | 1.148 | 0.152 | 0.411 | **Myh4** | myosin, heavy polypeptide 4, skeletal muscle |
| mCG18098.3 | 635108 | 0.816 | 0.166 | 0.431 | **Dzip1l** | DAZ interacting protein 1-like |
| mCG6928.1 | 766063 | 0.524 | 0.175 | 0.440 | **Optc** | opticin |
| mCG120470 | 594573 | 1.249 | 0.190 | 0.457 | **Hesx1** | homeo box gene expressed in ES cells |
| mCG11445.2 | 803233 | 0.830 | 0.335 | 0.602 | **Olig2** | oligodendrocyte transcription factor 2 |
| mCG18100.3 | 766206 | 1.095 | 0.366 | 0.626 | **D9Ertd280e** | DNA segment, Chr 9, ERATO Doi 280, expressed |
| mCG2198.2 | 320214 | 1.475 | 0.417 | 0.666 | | |
| mCG10171.1 | 743048 | 0.752 | 0.447 | 0.689 | **Spesp1** | sperm equatorial segment protein 1 |
| mCG21942.2 | 333605 | 0.934 | 0.472 | 0.710 | **Bnip3** | BCL2/adenovirus E1B 19kDa-interacting protein 1, NIP3 |
| mCG129238.2 | 649474 | 1.759 | 0.504 | 0.733 | **Vsnl1** | visinin-like 1 |
| mCG12846.2 | 926995 | 1.085 | 0.545 | 0.763 | **Zfp598** | zinc finger protein 598 |
| mCG52906.1 | 866474 | 1.264 | 0.571 | 0.780 | **na** | gene model 258, (NCBI) |
| mCG15433.2 | 527940 | 1.052 | 0.634 | 0.819 | | |

ii. In a comparison with mES differentiation, 21 (55.3%) were differentially expressed with 3 genes having a FDR < 0.05, namely Egr1, Gsto2 and Prkar1a.

**Table 6.15: 21 mouse genes from mES differentiation.**

| *Mouse Orthologs* | *ProbeID* | *FC (diff/ undiff)* | *p value* | *FDR* | *Gene Symbol* | *Gene_Name* |
|---|---|---|---|---|---|---|
| mCG123917.1 | 524988 | 2.082 | 7.77E-05 | 0.012 | **Egr1** | early growth response 1 |
| mCG142428.1 | 863193 | 0.238 | 0.002 | 0.035 | **Gsto2** | glutathione S-transferase omega 2 |
| mCG19628.2 | 848724 | 0.760 | 0.003 | 0.044 | **Prkar1a** | protein kinase, cAMP dependent regulatory, type I, alpha |
| mCG15433.2 | 527940 | 0.585 | 0.005 | 0.053 | | |
| mCG16156.2 | 880942 | 3.149 | 0.008 | 0.067 | **Trf** | transferrin |
| mCG1037923.1 | 897869 | 1.827 | 0.013 | 0.083 | **Tdgf1** | teratocarcinoma-derived growth factor |
| mCG18098.3 | 635108 | 0.581 | 0.018 | 0.099 | **Dzip1l** | DAZ interacting protein 1-like |
| mCG120470 | 594573 | 0.475 | 0.029 | 0.131 | **Hesx1** | homeo box gene expressed in ES cells |
| mCG10546.2 | 386024 | 2.547 | 0.055 | 0.188 | **D830007B15Rik** | RIKEN cDNA D830007B15 gene |
| mCG4792.2 | 885628 | 1.724 | 0.088 | 0.243 | **Ripk4** | receptor-interacting serine-threonine kinase 4 |
| mCG6928.1 | 766063 | 0.504 | 0.125 | 0.296 | **Optc** | opticin |
| mCG1039741.1 | 875245 | 1.385 | 0.148 | 0.326 | **Hbq1** | hemoglobin, theta 1 |
| mCG8982.1 | 450459 | 0.862 | 0.163 | 0.345 | **1810054D07Rik** | RIKEN cDNA 1810054D07 gene |
| mCG5518.1 | 463565 | 2.028 | 0.187 | 0.372 | **Fos** | FBJ osteosarcoma oncogene |
| mCG18100.3 | 766206 | 1.270 | 0.300 | 0.493 | **D9Ertd280e** | DNA segment, Chr 9, ERATO Doi 280, expressed |
| mCG11445.2 | 803233 | 1.373 | 0.417 | 0.598 | **Olig2** | oligodendrocyte transcription factor 2 |
| mCG140437 | 545994 | 1.175 | 0.481 | 0.651 | **Myh4** | myosin, heavy polypeptide 4, skeletal muscle |
| mCG21942.2 | 333605 | 0.748 | 0.511 | 0.676 | **Bnip3** | BCL2/adenovirus E1B 19kDa-interacting protein 1, NIP3 |
| mCG12846.2 | 926995 | 0.938 | 0.512 | 0.677 | **Zfp598** | zinc finger protein 598 |
| mCG10171.1 | 743048 | 1.203 | 0.649 | 0.782 | **Spesp1** | sperm equatorial segment protein 1 |
| mCG2198.2 | 320214 | 0.930 | 0.769 | 0.864 | | |

*Nullipotency in mouse and human cells*

Nullipotent cells in both models were treated as their differentiated counterparts and therefore followed the differentiation protocol. In the mouse model, only 4 genes were found to be significant, i.e. with a FDR < 0.05 while in the human model, only 12 genes were found. A comparison of these genes in both models was made.

Taking the 12 genes found in human nullipotent differentiation, only 8 had mouse orthologs. All 8 were not found to be differentially expressed in the mouse model even including genes with FDR < 0.25.

Analysis of the 4 genes found in mouse nullipotent differentiation is demonstrated in Table 6.16.

**Table 6.16: The four genes differentially expressed with FDR < 0.05 in mouse nullipotent EC cell differentiation and their equivalent in human nullipotent EC cell differentiation.**

| *Mouse Nullipotent EC cells* | | *Human Nullipotent EC cells* | | |
|---|---|---|---|---|
| **Gene Symbol** | **FC (diff/undiff)** | **Gene Symbol** | **FC (diff/undiff)** | **Gene_Name** |
| Oat | 0.634 | OAT | 0.954 | ornithine aminotransferase (gyrate atrophy) |
| Unknown | 1.920 | Not found | | |
| Gnpat | 0.582 | GNPAT | 0.947 | glyceronephosphate O-acyltransferase |
| Abi2 | 0.813 | Not found | | abl-interactor 2 |

Focusing on the comparisons of nullipotency and pluripotency of each model shows many similarities and some differences:

| | | |
|---|---|---|
| No. of genes upregulated in Nullipotency | 1575 (60.4%) | Human |
| | 1536 (46.9%) | Mouse |

**Table 6.17: Comparison of genes up and downregulated in mouse nulli versus pluripotent EC comparison with corresponding values in human data using human orthologs as the search criteria.**

| Mouse top 10 over and under expressed genes | | | Human comparison | | |
|---|---|---|---|---|---|
| *Gene_Symbol* | *FC (Nullipotent/Pluripotent)* | *FDR (BH)* | *Gene_Symbol* | *FC (Nullipotent/Pluripotent)* | *FDR* |
| Downregulated in Nulli versus Pluripotent | | | | | |
| **Dcn** | 0.001 | 3.09E-09 | DCN | 1.386 | 0.096 |
| **Hkdc1** | 0.003 | 3.80E-05 | Not found | | |
| ***Ctsh*** | *0.003* | *6.36E-06* | *CTSH* | *0.457* | *0.003* |
| **Pthr1** | 0.003 | 4.68E-06 | PTHR1 | 1.041 | 0.971 |
| **Prg1** | 0.004 | 8.59E-09 | Not found | | |
| **Tm4sf1** | 0.004 | 1.25E-08 | TM4SF1 | 0.859 | 0.838 |
| ***Fxyd3*** | *0.004* | *4.64E-06* | *FXYD3* | *7.549* | *0.001* |
| **1810036H07Rik** | 0.005 | 9.16E-09 | Not found | | |
| **Tnc** | 0.005 | 3.19E-09 | TNC | 0.467 | 0.235 |
| **Mglap** | 0.005 | 7.68E-07 | Not found | | |
| Upregulated in Nulli versus Pluripotent | | | | | |
| **Tsx** | 27.002 | 5.60E-07 | No ortholog | | |
| **Na** | 27.135 | 0.00041 | Not found | | |
| **Tm7sf4** | 27.490 | 2.12E-05 | Not found | | |
| **2610018G03Rik** | 27.785 | 0.000179 | No ortholog | | |
| **Aard** | 29.240 | 2.31E-07 | No ortholog | | |
| **Clec4a1** | 29.439 | 1.96E-06 | Not found | | |
| ***Crygd*** | *30.173* | *0.000577* | *CRYGD* | *18.243* | *0.0001* |
| **Rhbg** | 32.914 | 6.00E-05 | Not found | | |
| ***Hpgd*** | *43.036* | *0.000394* | *HPGD* | *15.458* | *0.002* |
| **Rtn4rl1** | 44.373 | 9.52E-06 | RTN4RL1 | 0.959 | 0.874806 |
| **Pramel6** | 468.271 | 2.39E-07 | No ortholog | | |

Four genes are highlighted as being significantly common to both mouse and human, namely Foxyd3, Ctsh, Crygd and Hpgd. Interestingly, all except Foxyd3 have similar fold changes with Foxyd3 downregulated in mouse nullipotent cells and upregulated in human nullipotent cells.

**Table 6.18: Comparison of genes up and downregulated in human nulli versus pluripotent EC comparison with corresponding values in mouse data using mouse orthologs as the search criteria.**

| Human top 10 up and downregulated genes in Nullipotent cells | | | Mouse Comparison | | |
|---|---|---|---|---|---|
| *Gene_Symbol* | *FC (Nulli/Pluri)* | *FDR (BH)* | *Gene_Symbol* | *FC (Nulli/Pluri)* | *FDR (BH)* |
| Upregulated in nulli versus pluripotent cells | | | | | |
| na | 1915.01287 | 1.33E-06 | Unknown | 2.109361736 | 1.45E-05 |
| LCP1 | 202.3533079 | 0.000168 | Not found | | |
| FLJ32942 | 140.1029794 | 0.000111 | 1700006H03Rik | 2.66042907 | 0.000474 |
| HORMAD1 | 83.66361744 | 9.10E-06 | Hormad1 | 1.795200775 | 0.00746 |
| CDH1 | 81.63263713 | 2.21E-05 | Cdh1 | 1.922877154 | 6.74E-07 |
| Unknonn | 77.31384672 | 4.04E-06 | No ortholog | | |
| CRI1 | 68.35200229 | 0.000926 | Cri1 | 0.402101335 | 0.002696 |
| C14orf29 | 61.58424487 | 1.98E-05 | No ortholog | | |
| THUMPD2 | 61.09912001 | 0.001038 | Unknown | 0.571005562 | 1.99E-05 |
| Unknown | 58.68502586 | 0.001612 | No ortholog | | |
| Downregulated genes in nulli versus pluripotent cells | | | | | |
| RARRES2 | 0.033985604 | 1.73E-06 | Rarres2 | 0.890660431 | 0.813556 |
| HMGA2 | 0.032867529 | 2.74E-06 | Hmga2 | 0.274234868 | 0.022573 |
| PIWIL2 | 0.030348779 | 3.95E-05 | Piwil2 | 0.84272351 | 0.417233 |
| | 0.029446425 | 0.00279 | 2810417H13Rik | 1.607900174 | 0.033057 |
| | 0.027724891 | 0.000101 | Unknown | 1.527167505 | 0.000116 |
| FEZ1 | 0.026288084 | 0.000111 | Fez1 | 0.352160838 | 4.79E-07 |
| ECAT8 | 0.024251234 | 6.48E-05 | No ortholog | | |
| LHX2 | 0.020506325 | 0.000106 | No ortholog | | |
| MYH6 | 0.012618216 | 1.94E-05 | No ortholog | | |
| | 0.00878449 | 7.84E-06 | BC010801 | 1.726769017 | 0.010152 |

More commonalities are present using the human data with 11 out of 20 genes common to both human and mouse nullipotent comparisons (highlighted in grey). Of the remaining nine genes, 6 had no corresponding mouse ortholog, with 1 not found and 2 with non-significant results, FDR > 0.05).

### 6.5.4 Validation of Microarray results

Validation of microarray results was performed using real-time PCR. The reasons for undertaking validation experiments are:

1. To verify that the observed changes are reproducible in a larger number of samples.

2. To verify that array results are not the result of problems inherent to the array technology.

Microarrays are known to be an excellent tool for initial target discovery but there are many sources of variability that may affect results including variability from laboratory to laboratory, user to user and platform to platform. Therefore, it is essential to use independent means to verify that the genes of interest are truly differentially expressed, and to what extent.

Validation was performed using 22 targets, details listed in Chapter 2, Table 2.12.

**Table 6.19: 22 TaqMan® gene targets for validation with fold change values and comparison with fold change values from microarray analysis for the differentially expressed genes upon differentiation of HP (Ntera2) and HC (2102Ep) cell lines.**

| | | HP diff v undiff | | | HC diff v undiff | | |
|---|---|---|---|---|---|---|---|
| | | TQ FC | Array FC | P value | TQ FC | Array FC | P value |
| 1 | AKT3 | 1.80 | 0.78 | 0.54 | 0.19 | 1.12 | 0.69 |
| 2 | CHN2 | 1314.72 | 4.05 | 0.23 | 1.63 | Not found | |
| 3 | DBC1 | 0.20 | 0.49 | 0.34 | 1.10 | 0.52 | 6.78E-03 |
| 4 | DLX5 | 125.05 | 2.90 | 0.19 | 10.91 | 0.69 | 0.24 |
| 5 | ENO3 | 6.62 | 5.39 | 0.05 | 0.66 | Not found | |
| 6 | FOS | 0.06 | 0.19 | 0.19 | 2.03 | 0.94 | 0.83 |
| 7 | FZD4 | 31.19 | 2.67 | 0.27 | 1.72 | 1.90 | 0.02 |
| 8 | FZD6 | 1.32 | 1.71 | 0.21 | 1.78 | Not found | |
| 9 | GPSM1 | 1.64 | 1.18 | 0.54 | 0.29 | 1.43 | 0.27 |
| 10 | HESX1 | 0.03 | 0.23 | 0.27 | 1.54 | 0.43 | 0.13 |
| 11 | JUN | 0.14 | 1.64 | 0.57 | Undet | 3.77 | 0.23 |
| 12 | NLE1 | 0.29 | 0.92 | 0.52 | 0.73 | 1.25 | 0.37 |
| 13 | OCT4 | 0.28 | 0.79 | 0.66 | 0.94 | 1.46 | 0.41 |
| 14 | PAZ6 | 85.85 | 5.50 | 0.29 | 0.96 | 1.71 | 0.35 |
| 15 | PLXNA2 | 704.57 | 6.28 | 0.09 | 0.47 | 1.20 | 0.25 |
| 16 | SCAP2 | 243.56 | 9.47 | 0.30 | 11.09 | 14.15 | 9.33E-05 |
| 17 | SOX9 | 0.82 | 0.54 | 0.007278 | 0.97 | 0.70 | 0.10 |
| 18 | SPRY1 | 0.59 | 0.32 | 0.21 | 1.93 | 0.52 | 0.16 |
| 19 | SPRY4 | 0.03 | 0.77 | 0.50 | 0.39 | 0.94 | 0.64 |
| 20 | ST7L | 0.91 | 3.15 | 0.001027 | 0.76 | 0.56 | 0.09 |
| 21 | TDGF1 | 0.02 | 0.19 | 0.36 | 0.57 | 0.65 | 0.10 |
| 22 | UTF1 | 0.05 | 0.99 | 0.99 | 2.70 | 0.78 | 0.10 |

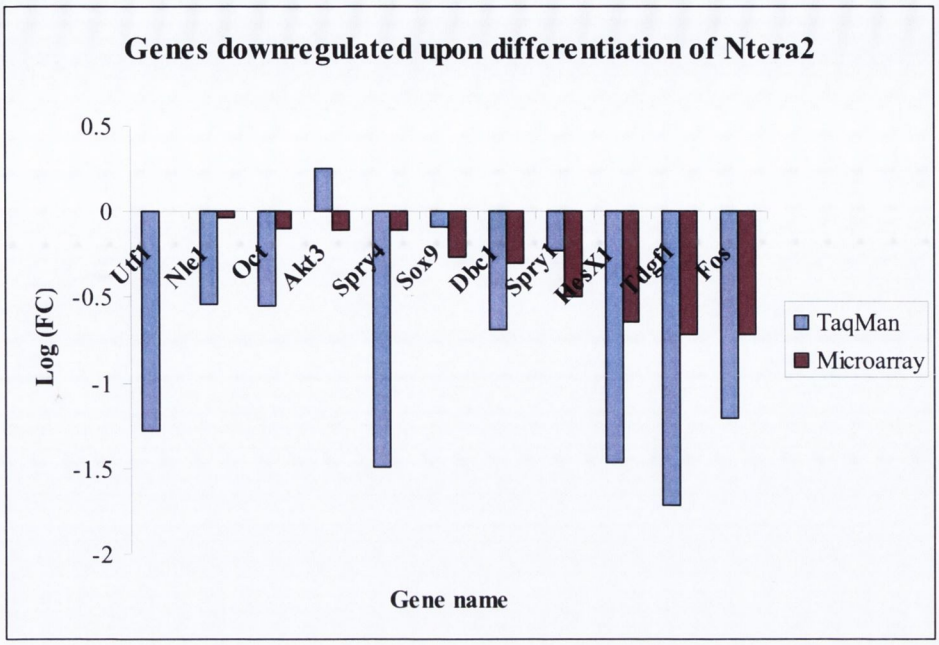



**Figure 6.9: Genes downregulated by microarray analysis upon differentiation of Ntera2 at day 3.**

Looking at the genes downregulated upon differentiation (11 in total), TaqMan® and microarray fold changes were broadly similar except with AKT3 which was downregulated by microarray analysis (fold change 0.78 or -1.28) and upregulated by TaqMan® analysis (fold change 1.80). However, in both cases, the fold change did not exceed 2 and was therefore most likely not significant.

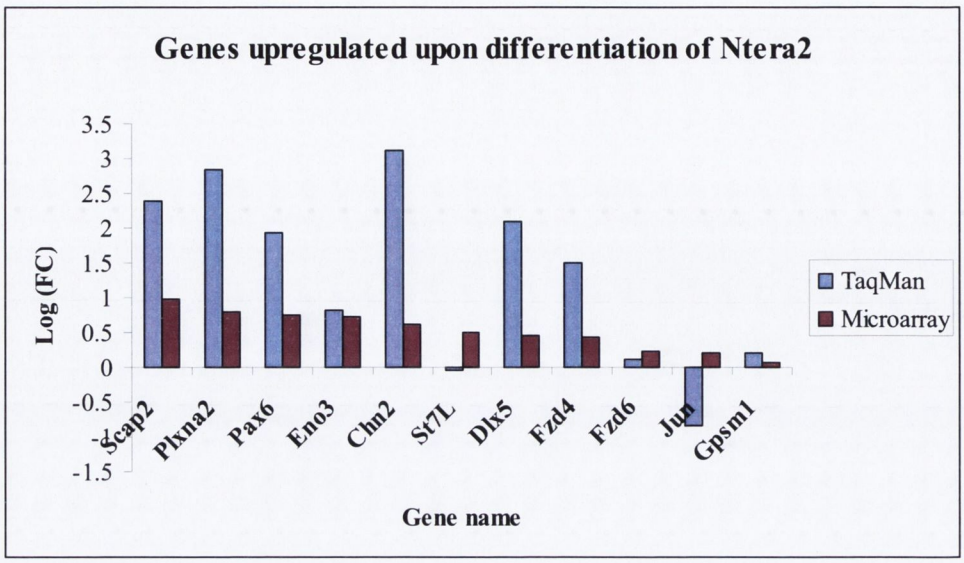



**Figure 6.10: Genes upregulated on microarray analysis upon differentiation of Ntera2 (TaqMan® = blue, Microarray = maroon).**

The upregulated genes upon differentiation of Ntera2 cells had broadly similar results with both TaqMan® and Microarray analysis except for JUN and ST7L. JUN was upregulated 1.64 fold with microarray analysis but downregulated with TaqMan® by 7 fold. ST7L was upregulated by 3.15 fold by microarray analysis and downregulated very marginally with TaqMan® analysis.

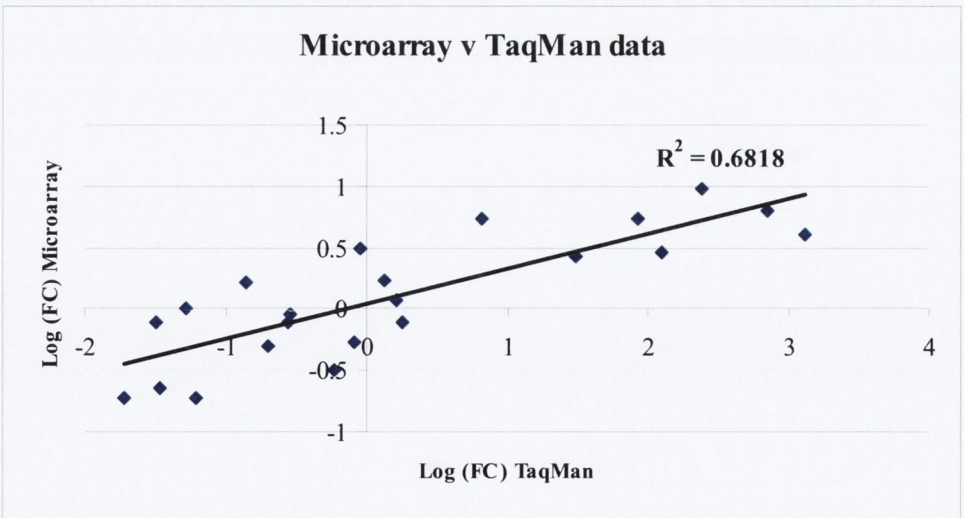



**Figure 6.11: Microarray versus TaqMan® data. Correlation coefficient = 0.83, Slope 0.28.**

The correlation coefficient for the data was 0.83 which shows good correlation between the TaqMan® and Microarray data; however it was not quite as good as the mouse data which had correlation coefficients of 0.97 and 0.86.

Using the validation data from the nullipotent HC cells (2102Ep), the correlation coefficient was less at 0.277 with some of the 22 genes not found on the array. However, in many cases, the p value for the fold change on the array was not significant and the fold changes using TaqMan® were $< 2$ or $> 0.5$.

## 6.6 Discussion

Using cRNA microarrays from Applied Biosystems with independent validation using quantitative real-time TaqMan® PCR, a comparison of pluripotent and nullipotent human embryonal carcinoma cell lines was performed as well as a comparison of undifferentiated and differentiated (at day 3) pluripotent and nullipotent cells in a similar manner to that performed in the mouse model as discussed in Chapters 4 and 5. The mechanisms governing differentiation of pluripotent stem cells such as Ntera2 are poorly understood (Andrews, 1998). Limited studies have been performed previously involving human embryonal carcinoma cells but rarely has such an early time point as 3 days been analysed particularly in the nullipotent cell line, which is unique to this study. This chapter also incorporates a comparison of differentiation in human pluripotent EC cells and hES cells, obtaining data on hES cells from the literature. However, many of the studies looking at hES cells involve a comparison of multiple hES cell lines to look for inherent differences between them with a number of studies analysing hES cell differentiation, often at time points of $> 8$ days or using mature differentiated tissues or other differentiated cell lines. Therefore, a limited comparison was performed with more emphasis and focus placed on the Ntera2 differentiation data as well as the comparison between Ntera2 and 2102Ep to determine the unique properties of 2102Ep that consistently promotes self-renewal and prevents differentiation.

*Differentiation in Ntera2 cell line*

A common theme in the literature illustrated in some of the examples above is to compare undifferentiated ES and or EC cells to mature differentiated tissues. One of the strengths of this study was using the differentiated counterparts of the undifferentiated cells for comparison at a very early time point in differentiation, i.e. 3 days. This allowed a much more specific analysis of the events occurring in differentiation, and more particularly a very early stage of differentiation. This analysis produced a list of 249 differentially expressed genes upon differentiation of Ntera2 cells with a p value $< 0.05$ and fold change $> 2$ or $< 0.05$ with the majority (69.5%) upregulated upon differentiation. A total of 76 genes were downregulated upon differentiation, thus enriched in undifferentiated pluripotent EC cells. Of these, 36 (47.4%) were unnamed or novel genes while the remaining 40 (52.6%) were known genes. Hierarchical clustering demonstrated that all the undifferentiated Ntera2 (HP) arrays cluster together while the three differentiated arrays clustered together as expected (Figure 6.5). While PANTHER functional analysis did not find any pathway, biological process or molecular function to be over represented in the differentially expressed data, some information regarding the pathways, biological processes and molecular functions involved in Ntera2 differentiation was extracted. From the data on pathways, the majority of genes were not classified into a pathway (84%) but individual pathways were looked at including angiogenesis, Wnt, Notch, and Hedgehog signalling pathways with various genes showing some changes. Figure 6.3 shows the biological processes with the highest number of genes contained within their group and includes metabolism of proteins and nucleotides as well as signal transduction. Transcription factors and nucleic acid binding were the molecular functions with the highest number of genes (Figure 6.4).

Looking at differentiation in Ntera2 cells (labelled HP cells), PNCK was the most highly over expressed gene upon differentiation with a fold change value of 12.3 and was also specifically involved in signal transduction, one of the biological processes containing a high number of differentially expressed genes. Perry Gardner et al, 2000, studied PNCK in the setting of mammary gland development as they set out to identify protein kinases that were expressed in murine mammary gland during development and in mammary tumour cell lines. PNCK is a serine/threonine kinase and a member of the CaM-dependent family of protein kinases. Calcium, a key intracellular signalling molecule, exerts its effect by binding to calmodulin and

activating CaM kinases (Perry Gardner et al., 2000). They found that PNCK was restricted to a subset of human breast tumour cell lines and was highly over expressed in a subset of primary human breast cancers. They concluded that PNCK may be expressed within a mammary epithelial cell type involved in differentiation as well as transformation. Their findings represent the first data implicating CaM kinase in mammary gland development and carcinogenesis.

The findings in this chapter of up regulation of PNCK in differentiated human pluripotent embryonal carcinoma cells would support their proposal that PNCK may be expressed by cells involved in differentiation as well as transformation to a cancer phenotype.
Why:

- Highly overexpressed in Ntera2 upon differentiation.
- Marginally overexpressed in SCC-PSA1 upon differentiation.
- Not differentially expressed in mES cells upon differentiation.

In mouse pluripotent teratocarcinoma cell lines, fold change of PNCK was 1.357, marginally upregulated upon differentiation with a p value > 0.5 and FDR > 0.5. In mouse embryonic stem cells, it was not present in the list of differentially expressed genes.

PNCK is predominantly expressed in the CNS and during development PNCK mRNA and protein are selectively expressed in the murine CNS throughout the mid gestation period. It has also been detected in other tissues including breast, uterus, brain, heart, and stomach. However, no functional studies of PNCK have revealed its role. A role in breast cancer was suggested in 2000 by Perry-Gardner as described above.

A study by Deb et al 2008, has identified PNCK as a component of a poorly understood mechanism of ligand-independent EGFR degradation (Deb et al., 2008). They propose that this observation raises the possibility that PNCK could be an endogenous protein inhibitor of EGFR overexpression in different cancers. Epidermal growth factor receptor (EGFR) and related members of the ErbB family are known to be amplified, overexpressed and activated in a variety of human cancers (Uberall et al., 2008). Their aberrant expression and activation results in growth signal autonomy and limitless replicative potential, the hallmarks of human cancer. Focusing on the origin and duration of signal at receptor level, promises much

therapeutic promise. Indeed, ligand-dependent EGFR degradation has been studied with much less attention on ligand-independent EGFR degradation as the mechanisms are not well understood.

Transmembrane channel-like 2 (TMC2) was the most downregulated gene upon Ntera2 differentiation with a fold change of 0.035 or -28.57. TMC2 belongs to the TMC family which have been named the transmembrane channel-like genes. The TMC gene family was discovered during the investigation of dominant and recessive sensorineural hearing loss (Keresztes et al., 2003). TMC1, a novel gene of unknown function, was shown to be mutated in some forms of hearing loss and has a similarity to TMC2, a gene with unknown function also, found on chromosome 20. It has been suggested that TMC proteins may function as ion channels, pumps or transporters (Kurima et al., 2003). TMC2 or other members of the TMC family have not been described previously in stem cells, embryonal carcinoma or other human or mouse tumours. Its precise role in stem cell regulation has yet to be determined.

The most downregulated signal transduction genes in differentiated Ntera2 cells (i.e. upregulated in undifferentiated Ntera2 cells) were AKAP5 (4.3 fold downregulated) and OPTC (5.2 fold downregulated. AKAP5 (alias AKAP79) belongs to a group of proteins known as A-kinase anchor proteins (AKAPs). It is thought to anchor protein kinase A (PKA) to cytoskeletal and/or organelle-associated proteins and to be involved in the β2-adrenergic (β2-AR) signalling pathway. Many AKAPs serve as organising centres for signal transduction by linking upstream signal generators to their downstream targets or by recruiting multiple signalling enzymes within signalling hubs (Gardner et al., 2006).

An interesting finding was the gene, SCAP2, which was upregulated upon differentiation of both Ntera2 and 2102Ep, indicating that this gene may play a significant role in carcinogenesis rather than in routine cell proliferation and differentiation. Indeed, it has recently been found to be significantly upregulated in pancreatic ductal adenocarcinoma and pancreatic intraepithelial neoplasia lesions (Harada et al., 2008). SCAP2 was originally described as a substrate for the src family kinases. The authors speculate that SCAP2 may control the growth and differentiation of pancreatic ductal adenocarcinoma cells and may be a future target of pancreatic cancer treatments. SCAP2 was not found to be differentially expressed in either mES or EC cells upon differentiation.

Of interest, chromosome 1 saw the highest number of genes differentially regulated in Ntera2 differentiation when compared with the other chromosomes. Chromosome 1 has previously been shown to be a common feature of paediatric solid tumours and while not specific for a given tumour, the frequent occurrence of chromosome 1 abnormalities may confer a clonal advantage in the development of cancer (Douglass et al., 1985). Abnormalities involving chromosome 12 are more frequently reported in association with germ cell tumours of both the testes and the ovary (Looijenga et al., 2003; Poulos et al., 2006) The discovery of involvement of chromosome 1 and the further pursuit of the associated genes may highlight further important gene targets involved in the development of cancer and in particular the cancer stem cell.

Skotheim et al, 2005, performed transcriptional profiling experiments on testicular germ cell tumours including seminoma and embryonal carcinoma using Agilent Human 1A oligomicroarrays (Skotheim et al., 2005). These results were compared with the in vitro profile of Ntera2 and 2102Ep. The aim of their study was to perform expression profiling on all the different histological types of testicular germ cell tumours to explore germ cell tumourigenesis and its transcriptional programs. Data from their analysis of Ntera2 and 2102Ep cell lines was compared with a comparison of embryonal carcinoma and differentiated tissues (normal testis, choriocarcinoma, yolk sac tumours, teratomas) and produced a list of 68 genes proposed to be characteristic of the undifferentiated embryonic stem cell-like or pluripotent state. Of these 68 genes, 45 (66.2%) were found to differentially expressed in the analysis of pluripotent EC cells performed in this chapter, however only 7 (10.3%) had a fold change > 2. The most highly differentially expressed genes of these were HESX1 and VSNL1 which were both > 4 fold downregulated upon differentiation.

Sperger et al, 2003, had a similar study using five hESC lines, EC cell lines (Ntera2, 2102Ep, NCCIT, TERA1, 833KE and 1777N) as well as normal testicular tissue, seminomas, embryonal carcinomas and yolk sac tumours (Sperger et al., 2003). Undifferentiated cells (ES and EC cells) were compared with differentiated tissues and tumours to identify genes specifically expressed at a higher level in pluripotent cell types. They also set out to prove that seminomas most closely resembled transformed primordial germ cells (PGCs) and EC cells most closely resembled transformed ICM or primitive ectoderm cells. Their comparison of

undifferentiated to differentiated cells involved undifferentiated ES cells being compared with differentiated tissues using somatic cell lines and yolk sac tumours.

*Differentiation in Ntera2 and hES cell lines*

Brandenberger et al, 2004, studied a comparison of hES cells to their differentiated counterparts (embryoid bodies at 8 days) and produced a list of 672 differentially expressed genes, a significant proportion of which were unknown, 38.6% (Brandenberger et al., 2004b). They concluded that a significant number of the differentially regulated genes were components of signalling pathways and transcriptional regulators and were likely to play key roles in hES cell growth and differentiation. In comparisons with other hES studies, they found a significant overlap with their own list of differentially expressed genes, specifically 76% and 71% with Ramalho-Santos et al and Ivanova et al (Ivanova et al., 2002; Ramalho-Santos et al., 2002). TDGF1, a member of the TGF-β family, was found to show the most significant differential expression (downregulated on differentiation) and is a regulator of the NODAL signalling. TDGF1 is known as a marker of undifferentiated hES cells (Bhattacharya et al., 2004; Brandenberger et al., 2004a). In my study of Ntera2 differentiation, TDGF1 was also found to be downregulated upon differentiation by 5.24 fold confirming its use as a marker of undifferentiated pluripotent stem cells. Other similarities are shown in Table 6.11. Interestingly, NANOG was not among their list of differentially expressed genes, and within my study, it was only marginally downregulated at 1.67 fold was not present in the list of 539 differentially expressed genes. The authors state that one of the strengths of their study was the ability to compare undifferentiated hES cells to their partially differentiated counterparts of the same genotype and passage, a similar strength to all experiments performed in this thesis. However, the work in this thesis intended to study an even earlier differentiation time point than most published results in the literature.

Bhattacharya et al produced two studies on hES cells, one of which presented a 92 gene signature of hES cells and the second study looked specifically at differentiation in hES cells. This 92 gene signature, genes over expressed by 3 fold or higher, constituted a molecular signature ("stemness") of hES cells and included several themes such as modulators of Wnt signalling, genes common to both mouse and human, zinc finger transcription repressors and cell cycle regulatory genes (Bhattacharya et al., 2004).

In their second study, Bhattacharya et al studied a hESC line and compared the undifferentiated state with EBs at 13 days differentiation using oligonucleotide glass arrays with 16,659 seventy-mer olignonucleotides produced in house (Bhattacharya et al., 2005). They produced a subset of 194 genes that were over expressed in EBs compared with hESCs with 157 of these known genes and 37 novel genes. Comparing their current data to a previously identified list of genes unique to hESCs, they found that the majority were downregulated upon differentiation as expected. They also noted that using different methods of gene expression detection such as MPSS and EST enumeration did produce some different results and therefore stressed the importance of using multiple methods to confirm gene expression.

In summary, a comparison of the Ntera2 differentiation with published data on hESC differentiation is difficult for many reasons:

- Different differentiation time points have been used, usually in excess of 8 days.
- Different array technologies, e.g. hESC analysis with Applied Biosystems array technology has not been performed to date.
- Different gene expression techniques such as MPSS.
- Many unknown genes are involved in hESC and hEC regulation.

While there are some similarities as highlighted, there are some differences which may be biological given the early differentiation time point used in this thesis.

*Genes Involved in Nullipotency*

There were two components to analysis of nullipotency in this chapter. The first was to look at the genes differentially regulated when nullipotent EC cells were stimulated to differentiate in the same manner as their pluripotent counterparts. This produced a list of 12 genes which were significant differentially expressed, 9 of which were known and 3 of which were novel genes. Four genes were downregulated and 8 were upregulated, the most highly upregulated of these was CYP26A1. The second approach was to compare the nullipotent EC cells in their undifferentiated state to pluripotent EC cells also in their undifferentiated state in an attempt to determine those genes unique to the pluripotent state and those to the nullipotent state. Genes

exclusive to the nullipotent state may provide novel insights into cancer stem cells with more sinister attributes. This produced a list of 2,607 differentially expressed genes (FDR < 0.05, fold change > 2 and < 0.5), 1,575 of which were upregulated in nullipotent EC cells and the remaining 1,031 were downregulated in pluripotent EC cells.

CYP26A1 was the most upregulated gene found in this study in 'differentiated' human nullipotent cells, i.e. downregulated in 'undifferentiated' nullipotent cells. It is also significantly (i.e. FDR < 0.05) differentially expressed in both the mouse and human comparisons of nullipotent and pluripotent cells being downregulated in nullipotent cells compared with their pluripotent counterparts (153 fold and 14 fold respectively). CYP26A1 is a metabolic enzyme, one of the cytochrome p450 enzymes, and has been shown to be expressed in some adult tissues, namely the liver and absent in others including skeletal muscle, kidney (Xi et al., 2008). The CYP26A1 gene is retinoic acid (RA) inducible and its transcriptional activation by RA acts to regulate RA levels. CYP26A1 has been shown by Bhattacharya et al, 2004, to be enriched in undifferentiated hES cells and include it in their 92 gene signature of hES cells (Bhattacharya et al., 2004). Langton et al studied its activity in mES cells and found that Cyp26a1 activity regulates intracellular RA levels, cell proliferation, and transcriptional regulation of primary RA target genes as well as differentiation to parietal endoderm. (Langton et al., 2008). They demonstrated that disruption of the Cyp26a1 gene resulted in increased intracellular RA and decreased sensitivity to growth inhibition by RA. Upregulation of CYP26A1 has been shown in cancer by Chang et al, 2008 in the setting of Barrett's carcinogenesis (Chang et al., 2008). Alterations in RA biosynthesis have been described in Barrett's oesophagus, the precursor of oesophageal adenocarcinoma so Chang and his group studied RA metabolism and the role of CYP26A1, a retinoic acid regulator that metabolises RA. They found that over expression of CYP26A1 in Barrett's oesophagus and oesophageal adenocarcinoma causes depletion of intracellular RA and thus drives the cell into a highly proliferative and invasive state with induction of other oncogenes. The finding in this thesis of CYP26A1 over expression in nullipotent EC cells being stimulated to differentiate is very interesting in light of this information as is it's over expression in pluripotent EC cells compared with nullipotent cells. It may either represent a mechanism of maintaining the cell against the effect of retinoic acid, the stimulus used for differentiation in these human EC experiments or support a role for CYP26A1 in the pluripotent phenotype as demonstrated by its over expression in pluripotent cells.

Josephson et al 2007 also looked at differences between 2102Ep and Ntera2 cell lines in their undifferentiated state using qRT-PCR and BeadArray experiments and found there were differences in expression of early germ cell markers with LEFTYB, STELLA, HOXB2 and WNT3 being expressed in 2102Ep with GAGE and PIWIL2 (later appearing germ cell markers) expressed in Ntera2 (Josephson et al., 2007). The data presented in this chapter corresponded somewhat to their findings with PIWIL2 one of the most highly downregulated genes in 2102Ep compared with Ntera2 (Table 6.9) and GAGE1 also being downregulated in 2102Ep compared with Ntera2 to a lesser degree. Also, LEFTY, WNT3 and HOXB2 were found to be over expressed in 2102Ep compared with Ntera2 in this chapter.

*Comparison of Human and mouse models*

Much has been written and published about the human and mouse ES transcriptomes to date, however, no detailed pairwise comparisons have been performed. Some studies have made attempts at a comparison and to date have found some similarities but also significant differences that need to be taken into consideration in the ongoing pursuit of the elusive definitive stem cell signature (Sato et al., 2003; Bhattacharya et al., 2004; Ginis et al., 2004; Wei et al., 2005). Comparisons are limited by the availability of cross-species and homologous arrays as well as the technology to analyse the vast quantities of genomic data obtained. The lack of published literature using Applied Biosystems array technology also limits a direct comparison of the data in this study to that in the literature. However, a comparison of differentiation in pluripotent hEC, mEC and mES cells was performed using the associated human and mouse orthologs but there were only a small number of overlapping genes. Taking the genes differentially expressed in the mouse model with associated human orthologs, only 2 genes were significant regulated in the human model, namely SEMA7A and LOXL1. Performing the reverse comparison using the differentially expressed genes in the human model resulted in no significant genes in mEC differentiation and 3 genes in mES differentiation, namely Egr1, Gsto2 and Prkar1a.

An identical comparison of nullipotent and pluripotent EC cells was performed in both the mouse and human model. Mouse and human orthologs were used to see if the most over and under expressed genes in nullipotent cells showed any similarities. Only genes with FDR $< 0.05$ and fold change $> 2$ or $< 0.5$ were considered significant. Four of 20 genes (Ctsh, Fxyd3,

Crygd and Hpgd) were found to be some of the most differentially expressed in mouse and subsequently in humans (see Table 6.17). Looking at the human data, 11 of 20 most highly differentially expressed genes in nullipotency were also significant in the mouse data (see Table 6.18). Of the remainder, many had no corresponding orthologs. These results are very interesting and given that concordance between mouse and human was low in the other comparisons performed, the high concordance demonstrated here may reflect the importance of this comparison in defining pluripotency.

## 6.7 Conclusion

Studies of human ES and EC cells focus much of their analysis of differentiation of stem cells on either mature differentiated tissues or differentiated cell lines to produce a molecular signature of stemness. This chapter presents analysis of Ntera2 differentiation at 3 days, a significantly earlier time point than most published studies, producing a list of 76 genes upregulated in undifferentiated Ntera2 cells. Within this list, 36 (47.4%) genes are novel while 40 (52.6%) genes are known. This list is proposed to represent a molecular signature of the pluripotent human embryonal carcinoma cell or cancer stemness.

Comparisons with human ES data has proven difficult given the wide variety of technologies and analysis methods used in the literature; however, it may be that the lack of similarities may be biological given the vast number of genes as yet unknown.

Analysis of nullipotency produced a list of 1,575 genes over expressed in the nullipotent state, a molecular signature of nullipotency as well as a list of 1,031 genes over expressed in the pluripotent state, a molecular signature of pluripotency. A limited comparison with the analogous mouse model demonstrated a high number of similarities which was reassuring as well as indicating that the increased similarities in this comparison may more accurately reflect genes important in defining pluripotency.

# CHAPTER 7 DISCUSSION

## 7.1 Overview

Stem cell biology has become an extremely popular area of research due to its implications in regenerative medicine and more recently in its role in the evolution of the cancer stem cell and carcinogenesis (Doss et al., 2004). This thesis uses stem cell biology to analyse the transcriptional profiles of a cancer stem cell using teratoma tumourigenesis as the model system. Differences between ES and EC cells (their malignant counterpart) should reflect the adaptation of EC cells to tumour growth while similarities should reflect genes necessary for the maintenance of the pluripotent or undifferentiated state thus producing a transcriptional signature of the cancer stem cell and the undifferentiated/pluripotent state. The mechanisms involved in the adaptation of EC cells to tumour growth lie at the heart of oncogenesis and involve the suppression of differentiation and subsequent promotion of self-renewal and survival. Outlining the detailed molecular networks underlying differentiation and self-renewal in these cells will ultimately enhance our understanding of the carcinogenic process and is key to the future management and treatment of cancer.

This thesis set out to investigate the molecular biology of cancer stem cell differentiation and self-renewal by obtaining transcriptional profiles of undifferentiated and differentiated stem cells and cancer stem cells using embryonal carcinoma or teratoma tumourigenesis in both mouse and human. While other groups have performed various comparisons of ES cells in mouse and human with a combination of differentiated tissues, this study was unique in looking at a very early time point of differentiation, i.e. 3 days, to try and ascertain the earliest molecular events occurring upon differentiation. As aberrant differentiation is thought to be at the heart of carcinogenesis, outlining the earliest molecular events might provide unique targets for future cancer therapies. As stem cells in other tissues share the properties of ES cells, this data using this model of embryonic stem cell and embryonal carcinoma is hoped to be widely applicable and not organ or tissue specific. The majority of stem cell groups are looking at stem cells and their use in regenerative medicine and thus looking at how differentiation may be controlled and directed with a view to producing replacement tissues rather than examining their role in carcinogenesis. A much smaller number of studies have looked at ES and EC cells in terms of gaining a cancer stem cell signature with this thesis

using a comparison of mouse ES, mouse EC and human EC cell differentiation to obtain a definitive cancer stem cell signature.

A further aim was to examine the fundamental differences between the nullipotent and pluripotent cell lines. Nullipotent embryonal carcinomas when xenografted produce undifferentiated embryonal carcinomas while pluripotent embryonal carcinoma cells produced teratomas, benign tumours composed of tissues of all three germ cell layers (Andrews et al., 1982; Duran et al., 2001). Poorly differentiated tumours, i.e. those that do not show evidence of differentiation, are known to have a much poorer prognosis than their more differentiated counterparts. Ben-Porath et al 2008 described this finding in human tumours and found that tumours enriched with genes associated with ES cell identity had a significantly poorer prognosis than tumours without this set of genes (Ben-Porath et al., 2008). The absence of differentiation capabilities and the molecular mechanism underlying this absence may be very revealing and uncover a fundamental cancer mechanism and as a result a useful source of future cancer targets. Thus, highlighting the fundamental molecular properties of nullipotent embryonal carcinoma cells might reveal some interesting findings in relation to carcinogenesis.

### 7.2 Mouse model

Mouse models are considered the model of choice in the study of human disease given that they share 99% of their genes and thus have played a fundamental role in biological research in multiple disciplines from development to cancer biology (Green et al., 2005; Abate-Shen, 2006; Degenhardt et al., 2006; Rosenthal et al., 2007). They have been used as model systems in the study of quite a number of cancers such as lung and pancreatic cancer (Shaw et al., 2005; Fomchenko et al., 2006; Olive et al., 2006). Indeed, the laboratory mouse (Mus musculus) has many advantages including ease of manipulating their genome, their small size, short lifespan (3 years), propensity to breed in captivity, extensive physiological and molecular similarities to humans and an entirely sequenced genome (Frese et al., 2007). They offer distinctive opportunities to investigate cancer mechanisms in genetically defined and environmentally controlled scenarios in the context of the tumour microenvironment. However, they should not be considered to replace the need for studies involving humans

given the many significant species differences between mouse and human, but to complement and augment studies of human cancer (Abate-Shen, 2006; Maser et al., 2007). Much of the initial work on mammalian embryogenesis and understanding the molecular events in cell differentiation and development was performed on mice, both in vivo and in vitro. The first cell lines to be developed for these purposes were mouse EC cell lines as far back as 1975 (Martin et al., 1975). Subsequently, mES cell lines were developed and are now widely used for various experimental purposes in cancer, regenerative and developmental biology.

### 7.2.1 Design and Implementation of Cell Culture Model and Time Course

Differentiation and self-renewal are the two main properties of stem cells and this study wanted to investigate these properties in relation to their influence on the carcinogenic process. To do this, a cell culture model had to be put in place. The main aim of the experimental work in Chapter 3 was to produce a cell culture model that would enable subsequent accurate transcriptome profiling of early differentiation and self-renewal in mES and EC cells. This involved an attempt to reproduce and validate other studies that had achieved feeder free cell culture conditions for the growth of mES and EC cells initially followed by the development of a reproducible differentiation time course. The development of feeder free cell culture conditions for both mES and hES cells would have a major impact on the sensitive array experiments performed by reducing contamination by a second cell population. In experiments on hES cells involving the production of different cell types for use in regenerative medicine such as replacement of cardiomyocytes following myocardial infarction or of replacement of pancreatic islet cells in Type 1 Diabetes Mellitus to re-establish insulin production, the removal of feeder fibroblasts would eliminate a major source of contamination. To this effect, experiments were set up to grow mouse EC cells in the presence and absence of LIF in the cell culture medium and spontaneous differentiation was allowed up to 3 days following removal of LIF. Cells grown without LIF were grown as per supplier's protocol, i.e. in the presence of a layer of mouse mitotically inactive fibroblasts. Following harvesting of cells at day 0 and day 3 differentiated and undifferentiated, RT-PCR was performed on a panel of markers to demonstrate whether differentiation had occurred, in combination with careful morphological examination. Ultimately, RT-PCR proved time consuming and laborious with equivocal results and therefore quantitative real-time RT-PCR

was subsequently performed to more accurately observe the expression of the panel of differentiation and self-renewal markers. Despite numerous attempts using several time course experiments, this study was unable to successfully achieve and reproduce feeder-free cell culture conditions as other authors have done previously. Mouse ES and EC cells are recognised to be more difficult to manage in cell culture experiments given their ability to spontaneously differentiate. Therefore, careful observation and optimal cell culture conditions are required at all times.

This study as described in Chapter 3 was unable to reproducibly achieve feeder-free cell culture as some other authors were, feeder cell layers were used in experiments for subsequent array analysis. This proved to be easier to manage and improved reproducibility dramatically. Looking at the qRT-PCR results for the time course experiments, changes in gene expression were noted at 3 days most notably with AFP. AFP showed a dramatic increase in expression following differentiation at 3 days in SCC-PSA1 cells (i.e. pluripotent teratocarcinoma cells). These results together with the observation of subtle morphological changes and results from other studies indicating that a large proportion of gene changes in differentiation are occurring in the first 2 weeks of differentiation, expression array analysis of cells from the cell culture model was performed at day 0 and day 3 of differentiation. The aim was to obtain transcriptome profiles of early differentiation and self-renewal of mES and EC cells using differentiated cells at day 3 post spontaneous differentiation thus examining the two major properties of stem cells and cancer stem cells with a view to highlighting potential sources of future cancer therapies.

### 7.2.2 The Transcriptome Profile of the Cancer Stem Cell

Chapter 4 of this thesis provides a direct comparison of mouse ES cells in their undifferentiated state with their malignant counterpart, mouse pluripotent teratocarcinoma cells with a view to outlining the fundamental differences between the cancer stem cell and its normal or 'benign' equivalent. An additional comparison with a nullipotent teratocarcinoma cell line was made both to incorporate as many differences and similarities between mES cells and EC cells as possible as well an analysis of the unique properties of nullipotent cells that maintain them in a constant self-renewing state, a property vital to cancer cells. Following

expression array analysis and subsequent data analysis to compare mES cells to EC cells in their resting states, some interesting findings were uncovered. A list of 1,170 differentially expressed genes was obtained from the comparison of mES cells to both pluri- and nullipotent EC cells. Using PANTHER for functional analysis and thus examining the various pathways, biological processes and molecular functions that were over-represented in this list of 1,170 genes, the oxidative stress pathway figured strongly as the most over-represented pathway identified. The oxidative stress pathway involves the control of mechanisms to protect the cell from ageing which would be important for stem cells to retain their permanent status within the body. This was first proposed by Ramalho-Santos et al where the authors performed a comparison of ES cells, haemopoietic stem cells, neural stem cells and differentiated cells from the bone marrow and lateral ventricles of the brain (Ramalho-Santos et al., 2002). They found that stem cells appeared to have characteristics of cells under stress.

Dusps, a subclass of protein tyrosine phosphatases uniquely able to hydrolyse the phosphate ester bond on both a tyrosine and a threonine or serine residue on the same protein, are genes involved in the oxidative stress response which were found to be differentially regulated in this array comparison, not described previously in stem cells. Dusps have crucial roles in intracellular signal transduction pathways and most notably regulate mitogen-activated protein kinase (MAPK) signalling pathways and cell cycle progression. In this array study, Dusp9 was significantly over expressed in malignant compared with normal stem cells. There is very little written in the literature regarding DUSP9 expression in human tissues. Christie et al, 2005, demonstrated that Dusp9 was expressed in mES cells and its removal resulted in embryonic lethality due to placental insufficiency rather than being absolutely necessary for embryonic development (Christie et al., 2005). This gene has not previously been described in malignancy. Another of the Dusp family, namely Dusp26 has been described as over expressed in anaplastic thyroid cancer (ATC) cell lines and tumour samples by Yu et al, 2006 (Yu et al., 2006). They found that Dusp26 promoted ATC cell growth by inhibiting p38 mediated apoptosis. They propose that it may be a useful diagnostic marker and potential therapeutic target for this very aggressive form of thyroid cancer. Interestingly, it was found to be downregulated in malignant cells in this study. Other Dusps found to be differentially expressed between the malignant and benign groups were Dusp6 (downregulated in malignant group) and Dusp16 (upregulated in the malignant group). There are reports of investigating the

targeting of DUSPs by inhibitory agents in the treatment of both cancer and Alzheimer's disease (Lyon et al., 2002).

A notable finding in this chapter was that of genes within the extracellular matrix group being significantly over-represented in the group of differentially expressed genes between mES and EC cells. The suggestion of a cancer stem cell niche is relatively new and more evidence is awaited to advance and further understand this concept (Vermeulen et al., 2008). The role of the stem cell niche in the regulation of stem cell function has been described and found to be vital to the management of stem cell differentiation and self-renewal (Adams et al., 2007). The findings in this chapter would appear to support the significant role of the stem cell niche and suggest a number of extracellular matrix proteins that may be useful future targets to base further research on in the future including Decorin and Tenascin C. These genes were found to markedly downregulated in nulli and pluripotent teratocarcinoma cells i.e. malignant stem cells, compared with mES cells i.e. normal stem cells. Decorin is a member of the family of small leucine-rich proteoglycans involved in a number of cellular processes. Interestingly, reduced Decorin expression has previously been described in a number of tumours, namely hepatocellular carcinoma, pancreatic carcinoma and some lung and ovarian tumours, thus suggesting that this gene plays a significant role in the carcinogenic process and adding further weight to the importance of the role of the extracellular matrix in cancer (Koninger et al., 2004; Seidler et al., 2006). Reed et al, 2005, investigated the role of Decorin in suppressing primary breast carcinomas and found that Decorin resulted in marked growth suppression both *in vitro* and *in vivo* with reduced primary tumour growth and a reduction in observed metastases (Reed et al., 2005).

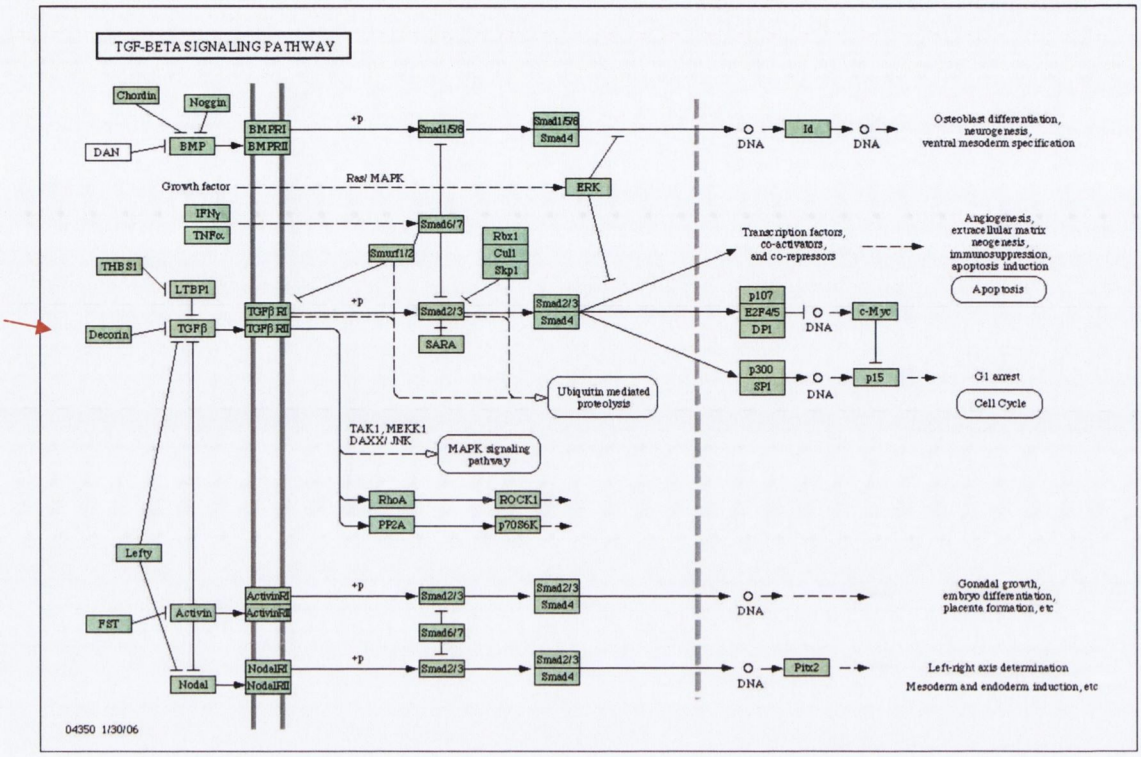



**Figure 7.1: TGF-β signalling pathway illustrating the role of Decorin (red arrow). Taken from KEGG, the Kyoto encyclopedia of Genes and Genomes, www.genome.ad.jp/kegg/, an initiative by the Japanese human genome programme, 1995 (Kanehisa et al., 2000).**

Decorin appears to have growth suppressive properties and in general is found in quiescent cells with absent expression in transformed cells (Seidler et al., 2006). This finding is supported by the findings in this thesis as Decorin in upregulated in normal or 'benign' stem cells and downregulated in the malignant group. Decorin may offer hope for future cancer treatments and Seidler et al, 2005, have been investigating possible methods of delivery of Decorin to inhibit *in vivo* tumour growth. Given the findings in this thesis, Decorin may be more efficient at targeting the cancer stem cells within tumours as the cancer stem cells display similar Decorin expression profiles to many other tumour cell types.

In summary, this chapter has produced a unique transcriptome profile of the cancer stem cell using a mouse model of ES and both pluri- and nullipotent EC cells which has not previously been performed. The profile consists of a list of 1,170 differentially expressed genes from which functional analysis and highly significant genes were obtained.

### 7.2.3 The Transcriptome Profile of the Differentiated and Self-Renewal States in Cancer Stem cells

While chapter 4 focused on the undifferentiated cancer stem cells, chapter five focused on differentiation and the genes differentially expressed upon differentiation in both mES cells and EC cells using an early differentiation time point of three days. As differentiation lies at the core of oncogenesis, the molecular mechanisms governing its regulation are crucial to understanding cancer and therefore this thesis aimed to focus on one of the earliest time points at which molecular changes are occurring. Previous studies have been undertaken using time points from 7 days after differentiation or comparisons of mES cells and mature differentiated tissues (Ivanova et al., 2002; Ramalho-Santos et al., 2002; Sperger et al., 2003; Choi et al., 2005; Heo et al., 2005). These studies have shown that the majority of the transcriptional changes are occurring within the first 7 days of differentiation (Heo et al., 2005). Therefore, this thesis has produced unique profiles of 'benign' and 'malignant' differentiation by using an early differentiation time point, i.e. 3 days, in both mES and EC cells (both pluri- and nullipotent). Looking at mES differentiation, 592 genes (FDR $< 0.05$ and fold change $\geq 2$ or $< 0.5$) were differentially expressed upon differentiation at 3 days with the majority, 443 (75.8%) downregulated. Pluripotent EC differentiation at 3 days produced 224 genes that were differentially regulated again with the majority of genes downregulated, 168 (75%). Nullipotent EC differentiation produced only 4 differentially expressed genes when only those with a FDR $< 0.05$ were considered, namely Oat, Gnpat, Abi2 and one unknown gene. The fold change, however, for all 4 genes was less than 2, with the unknown gene upregulated and the other 3 genes downregulated upon differentiation. Other studies of stem cell differentiation also show that the majority of differentially expressed genes are downregulated (Heo et al., 2005). A further analysis of these three groups of differentially expressed genes using Spotfire® data analysis software allowed the determination of those genes unique to mES, pluripotent EC and nullipotent EC differentiation. This comparison, not previously performed or obtained, produced twenty genes that were unique to pluripotent EC differentiation, 46 to mES differentiation and none identified as unique to nullipotent EC differentiation. These 66 genes are of immense interest in determining the genes at the centre of stemness with the 20 genes unique to pluripotent EC differentiation of the most relevance in finding a suitable target to eliminate the offending cancer stem cell population within tumours.

#### *7.2.3.1 Genes Common to ES and EC Cell Differentiation*

Two of the most significantly downregulated genes upon differentiation common to both mES and EC cells were Ccl5 and Oasl2. Ccl5 belongs to the chemokine family, a group of molecules involved in the regulation of the inflammatory response as well as a potential source of growth factors for tumour cells (Gerard et al., 2001). Oasl2 belongs to a family of interferon-induced antiviral proteins found in both humans and mice known as the 2'-5'-oligoadenylate synthetases (Eskildsen et al., 2003). These proteins are important components of the mammalian innate immune system. The differential expression of a chemokine and interferon related gene in early stem cell differentiation correlates with the PANTHER data which found that the biological processes of immunity and defense and interferon related immunity were over-represented in the data obtained from mES and EC differentiation. Also, chemokines were one of the molecular function groups to be over represented in mouse EC differentiation data. Interestingly, the chemokine Ccl5 has been described in the literature in various human cancers including lung, ovarian and gastric cancers (Borczuk et al., 2008; Sugasawa et al., 2008). Tumours with over expression of Ccl5 have been shown to be more likely to progress and metastasise and are thus associated with a poorer prognosis. The finding of downregulation of Ccl5 in stem cell differentiation indicates that that Ccl5 may be a useful target in the successful elimination of the 'cancer stem cell' component in tumours as the inactivation of Ccl5 would trigger differentiation rather than continuous maintenance of the tumour cell. No studies to date have described Oasl2 expression in tumours and its further characterisation together with Ccl5 may reveal potential useful cancer targets.

Similarities with other studies of mES cell profiling include finding that transcription factors, signalling molecules and extracellular matrix proteins are over represented in mES differentiation. Focusing on the group of signalling molecules differentially expressed upon stem cell differentiation revealed some fascinating findings as well as some differences between mES and EC differentiation. A significant risk in the treatment and successful elimination of cancer stem cells is the coincidental elimination of normal stem cells which may be of crucial importance in some instances. Therefore, highlighting the differences between mES differentiation and EC differentiation may provide more useful potential therapeutic targets. Fibulins (Fblns) are a family of widely expressed extracellular matrix proteins that belong to the signalling molecule group as well as the extracellular matrix group

in this data. Fbln2 and 5 were both shown to be differentially expressed in stem cell differentiation with Fbln2 found in both mES and EC differentiation while Fbln5 was unique to EC differentiation, downregulated by 6.5 fold upon differentiation. Reduced Fbln5 expression has been documented in the literature to be associated with tumour progression and proliferation (Schiemann et al., 2002; Albig et al., 2006). Fbln5 has also been shown to play a role in angiogenesis with an increase in angiogenesis being observed in the absence of Fbln5 (Sullivan et al., 2007). Thus Fbln5 is thought to suppress tumour growth by the inhibition of angiogenesis and inhibition of epithelial cell proliferation (Schiemann et al., 2002; Sullivan et al., 2007). Fbln2 has been shown in other studies to be downregulated in mES differentiation (Palmqvist et al., 2005). A microarray study of human colorectal carcinoma by Wiese et al 2007 demonstrated down regulation of Fbln2 in microdissected tumour samples with Fbln2 included in their 149 gene signature for colorectal carcinoma (Wiese et al., 2007). They suggest that Fbln2 belongs to the stromal cell or extracellular matrix component of colorectal carcinoma. While both Fbln2 and 5 appear to play a significant role in the differentiation and carcinogenic process, the data from this thesis suggests that Fbln5 may be a better choice for further study given that it is unique to the malignant cancer stem cell. This would allow targeting of the 'cancer stem cell' without affecting the native stem cells and thus elimination any potential harmful side effects of a cancer therapy.

#### *7.2.3.2 Genes Unique to EC Differentiation*

Focusing on only those genes with a FDR of $< 0.05$ and those with a fold change of $> 2$ results in only 20 genes unique to pluripotent EC differentiation with 11 of these recognised and the remaining 9 being unnamed. The two genes downregulated upon differentiation were Adamts5 (4 fold by microarray analysis, 7 fold by TaqMan® analysis) and Tnfrsf9 (2.3 fold by microarray analysis, 4.3 fold by TaqMan® analysis). Adamts5 was one of the first aggrecanases discovered. Aggrecanases are proteinases that cleave the Glu-Ala bond of the aggrecan core protein and are multidomain metalloproteinases belonging to the Adamts (adamalysin with thrombospondin type 1 motifs) family (Gendron et al., 2007). Aggrecanases are responsible for cartilage degeneration in osteoarthritis (Kevorkian et al., 2004; Gendron et al., 2007). Gendron et al 2007 have shown that Adamts5 is a major aggrecanolytic metalloproteinase secreted by the cell into the extracellular matrix (Gendron et al., 2007).

Much of the published literature on Adamts5 is in relation to its destructive role in osteoarthritis with inhibition of Adamts5 and other proteinases being investigated as a mode of treatment in osteoarthritis (Nagase et al., 2003; Kevorkian et al., 2004). However, there are some studies describing its expression in human glioblastomas, highly malignant brain tumours (Held-Feindt et al., 2005; Nakada et al., 2005). Glioblastomas are highly invasive and rely on increased expression of proteinases to digest the extracellular matrix with Adamts5 being one of these proteinases. The significance of Adamts5 downregulation upon differentiation in EC differentiation in this thesis is uncertain given the lack of published literature and has not been described previously. Its role as an extracellular matrix protein again highlights the importance of the extracellular matrix in cell proliferation and self-renewal.

The tumour necrosis factor (TNF) superfamily is a group of cytokines with an important role in immunity, inflammation, differentiation, apoptosis and control of cell proliferation with TNFalpha the founding member (Shen et al., 2006). TNF family members exert their biological effects through the use of the TNF receptor (TNFR) superfamily of cell surface receptors. TNFR superfamily, member 9 (Tnfrsf9) (aliases CD137 and 4-1BB) has been shown to be expressed by various immune cells including activated T cells, natural killer (NK) cells, monocytes, neutrophils and dendritic cells and plays a role in the amplification of T cell immunity (Lee et al., 2005). Agents against Tnfrsf9 or CD137 are being investigated for their use in the treatment of Rheumatoid arthritis (Foell et al., 2004) with other tumour necrosis factor antagonists being widely used in the treatment of immune-mediated inflammatory diseases (Wong et al., 2008). Martinet et al 2002 describe the role of 4-1BB or Tnfrsf9 in the management of metastatic breast cancer (Martinet et al., 2002). As T cell immunity is important in maintaining a tumour-free status, its amplification using anti-4-1BB antibodies as described by Martinet et al caused regression of metastatic breast tumours within the liver. A similar finding using 4-1BB mediated immunity was demonstrated in metastatic colon carcinoma (Chen et al., 2000). Other studies have also demonstrated the anti-tumour immune response mediated by 4-1BB and its potential role in the treatment of various tumours (Kim et al., 2001). It would appear from the published literature on Tnfrsf9/CD137/4-1BB that it plays an important role in the immune response both in inflammatory-mediated conditions and cancer. Its role in stem cells and stem cell differentiation has not previously been described in depth. From the studies involving cancer and Tnfrs9/4-1BB in the literature, it seems to play a

role in preventing metastases and its upregulation in undifferentiated EC cells or the 'malignant' stem cell described in this thesis is in keeping with this role and warrants further investigation. Thus, the microarray analysis of mouse EC differentiation in this thesis has revealed two interesting genes downregulated upon differentiation, unique to EC cells or the malignant stem cell, which may be potential targets for cancer stem cell therapies without the detrimental effects on normal stem cells.

Looking at some of the 9 genes upregulated in and unique to mouse pluripotent EC differentiation, Tex12 expression appears to be restricted to male germ cells (Pohlers et al., 2005; Hamer et al., 2006). This would indicate that Tex12 reflects the origin of the embryonal carcinoma itself and indeed the particular mouse embryonal carcinoma cell line, SCC-PSA1, used in these experiments was derived from a testicular embryonal carcinoma.

The insulin-like growth (IGF) signalling pathway has been found to be activated in numerous tumour types playing a key role in human malignancies through regulation of apoptosis and tumour cell proliferation. The IGFs are regulated by IGF binding proteins (Igfbps). Igfbp5 is one of a family of six structurally related proteins with affinity for Igf-I and Igf-II (Graham et al., 2007). These proteins are all found in the extracellular matrix. Igfbps either inhibit or enhance the action of Igfs. In a microarray comparison of squamous cell carcinomas of the cervix and normal cervical squamous epithelium, Miyatake et al, 2007, found IGFBP5 to be significantly downregulated in the squamous cell carcinoma group (Miyatake et al., 2007). Johnson et al, 2006, report that IGFBP5 may play a significant role in the malignant transformation of pancreatic epithelial cells into pancreatic adenocarcinoma (Johnson et al., 2006). Another group described downregulation of IGFBP5 to be responsible for tumour progression in neuroblastoma cell lines (Tanno et al., 2006). IGFBP5 has also been shown to be expressed in ovarian carcinoma, however, not in all histological types (Wang et al., 2006a). Wang et al 2006 demonstrated over expression of IGFBP5 in high grade serous carcinomas of ovary and not in low grade serous carcinomas, mucinous or clear cell carcinomas (Wang et al., 2006a). They accept that the role of IGFBP5 in malignancy is not yet clear and that their study only provides circumstantial evidence of a role for IGFBP5 in the malignant process. The finding in this thesis of upregulation of Igfbp5 in differentiated mouse EC cells and downregulation in undifferentiated EC cells together with differing results in the literature

would suggest either Igfbp5 expression is specific to a tumour type or its role in malignancy is not yet apparent.

Down syndrome cell adhesion molecule (Dscam) found to be upregulated in differentiated pluripotent EC cells only is a member of the immunoglobulin superfamily that maps to the chromosome at 21q22.2-22.3 (Li et al., 2004). It is mainly expressed in the brain and is thought to play a role in the formation of neural networks. Dscam is a type 1 transmembrane protein that likely functions as a cell surface receptor mediating the growth of axons. Its expression in differentiated EC cells in this study may reflect the occurrence of neural differentiation rather than any significant carcinogenic role.

Therefore, looking at the genes differentially expressed in early differentiation unique to pluripotent EC cells, the genes downregulated on differentiation would appear to be more significant and worth pursuing in relation to future therapies rather than the upregulated genes. The set of upregulated genes including Dscam, Igfbp5 and Tex12 described above appear to have more wide ranging roles and may be either specific to the individual tumour type or simply indicate that differentiation has occurred. Adamts5 and Tnfrsf9 show more promise as novel cancer targets as they have increased expression in the self-renewal state in cancer stem cells and were not shown to be differentially regulated in the normal stem cell i.e. mES cells in this study.

#### *7.2.3.3 Genes Involved in Nullipotency*

Nullipotent EC differentiation demonstrated that 4 genes were differentially expressed upon differentiation, namely Oat, Gnpat, Abi2 and one unnamed gene. Nullipotent EC cells are known for their inability to differentiate and the differentiation described here is based simply on the fact that the differentiation protocol for pluripotent EC cells was applied to the nullipotent group. The 4 genes were found however to have a fold change of $< 2$ and therefore most likely of no significance.

Of potentially more importance are the genes differentially expressed in nullipotent compared with pluripotent embryonal carcinoma cells. This analysis produced a list of 3,275 genes when restricted to genes with FDR $< 0.01$ and fold change $> 2$ and $< 0.5$. Of these genes, 46.9%

were upregulated and 53.1% were downregulated in nullipotent compared with pluripotent cells with Decorin and Tenascin C again featuring significantly. As discussed previously, in a comparison of mES versus EC cells (both nulli and pluripotent), Decorin and Tenascin C were downregulated in both types of EC cells while in a comparison of nulli- and pluripotent cells, both genes were upregulated in pluripotent EC cells. This data suggests that both Decorin and Tenascin C play a significant role in tumourigenesis and cancer stem cells in particular being not only downregulated in the malignant group but also downregulated in the nullipotent cell line (a cell line that maintains a constant state of self-renewal) compared with the pluripotent cell line.

Taking the three main regulators of the self-renewing state, namely Nanog, Oct4 and Sox2, all 3 are slightly upregulated in nullipotent EC cells compared with their pluripotent counterparts with Nanog 3 fold upregulated in nullipotent cells (see Table 5.16). These results might be expected as nullipotent cells try to maintain a continuous self-renewing state. This data would also appear to confirm that Nanog is indeed an important gene in the self-renewal state as described in the literature (Torres et al., 2008).

The most highly upregulated gene in nullipotent compared with pluripotent cells was Pramel6. The PRAME gene family are a group of tumour antigens shown to be over expressed in a wide variety of cancers including melanomas, non-small cell lung carcinomas, breast carcinoma, renal cell carcinoma and Hodgkin lymphoma (Epping et al., 2006). A related group of genes are the PRAME-like (Pramel) genes which have been demonstrated in the human and mouse genome. Pramel6 expression has been demonstrated along with Pramel7 in embryonic stem cells in mice. Its dramatic up regulation in nullipotent compared with pluripotent embryonal carcinoma cells in this study by microarray analysis and confirmed by TaqMan® analysis is very interesting. Does Pramel6 have a role in the maintenance of the self-renewing state, thus conferring a survival advantage to the nullipotent embryonal carcinoma cells? The data would seem to suggest so and given that little is known and written about Pramel6 in the literature; its further elucidation would be worth pursuing to determine its precise role in the maintenance of the cancer stem cell.

Looking at the functional analysis of the differentially expressed genes from PANTHER, signalling, select regulatory and extracellular matrix proteins are important as they were in

comparisons of differentiated and undifferentiated pluripotent cells. Similarly, comparable biological processes are involved, i.e. developmental processes, cell proliferation and differentiation. Angiogenesis, an essential pathway in carcinogenesis as well as normal vascular development was a significant pathway in the nulli versus pluripotent comparison as well as in pluripotent EC differentiation confirming that angiogenesis is important in carcinogenic process and may have a vital function in the maintenance of the cancer stem cell.

### 7.3 Human model

The replication of the mouse model outlined in chapters 3, 4 and 5 in a human system would require the use of human ES cells for comparison with human pluri- and nullipotent embryonal carcinoma cell lines. While the two tumour cell lines are easily obtainable and indeed the human EC cell lines needed, Ntera2 and 2102Ep, were received as a gift in our laboratory from Professor Peter Andrews of the University of Sheffield, the use of human ES cells for experimental purposes is not ethically permissible in Ireland. To this end, the human EC cells were grown as per protocol outlined in Chapter 2 in a similar time course to their murine counterparts. Data obtained was compared with published literature on human ES cells and other work on human EC cells. The direct comparison of an embryonic stem cell and embryonal carcinoma cell performed in the mouse model was more difficult to achieve as all the array data within this thesis was performed using the Applied Biosystems array platform while much of the data on hES cells in the literature is performed using other array platforms including Affymetrix and various in house array platforms. Of particular interest from the human EC data was the comparison of nulli- and pluripotent cells to determine what allows nullipotent cells to remain incapable of differentiation.

The comparison of undifferentiated pluripotent EC cells to differentiated cells at 3 days differentiation produced a list of 539 genes when genes with a p value $< 0.05$ and fold change $> 2$ and $< 0.5$ were considered. No differentially expressed genes were listed using a FDR $<$ 0.05 or 0.25 so for the purposes of this comparison the p value was considered an adequate alternative but less than ideal. This lack of highly significant gene changes may be due to the use of a very early time point in differentiation, thus producing only minor gene changes. Of the list of 539 genes, the majority or 68.1% were upregulated upon differentiation. The fold

changes involved were also quite low relative to the fold changes seen in differentiation in mouse EC cell differentiation, again this may reflect the lack of significant gene changes occurring at this early time point of 3 days.

There were two components to the analysis of nullipotency in this model, the first examined differentially expressed genes between nullipotent EC cells and nullipotent EC cells which were stimulated to differentiate in a similar manner to their pluripotent counterparts. A total of 12 genes were found to be significantly differentially expressed when this analysis was performed with 4 downregulated and 8 upregulated in the 'differentiated' nullipotent cells. Of these, 9 known genes and 3 novel genes were identified. The most highly upregulated gene was CYP26A1 which is a metabolic enzyme belonging to the group of cytochrome p450 enzymes. It has been shown to be expressed in some adult tissues, namely the liver and absent in others including skeletal muscle, kidney (Xi et al., 2008). Bhattacharya et al included it in their 92 gene signature of hES cells, demonstrating it to be enriched in undifferentiated hES cells (Bhattacharya et al., 2004). Another study by Langton et al established its role in mES cells and found that Cyp26a1 was crucial for mES cell proliferation and differentiation (Langton et al., 2008). One recent study was found in the literature illustrating CYP26A1 in the cancer setting. Chang et al, 2008, found over expression of CYP26A1 in Barrett's oesophagus, the precursor of oesophageal adenocarcinoma (Chang et al., 2008). The functional effects of this were increased cell proliferation and invasiveness. CYP26A1 was found to be significantly (i.e. FDR $< 0.05$) differentially expressed in both the mouse and human comparisons of nullipotent and pluripotent cells in this thesis, being downregulated in nullipotent cells compared with their pluripotent counterparts in both (153 fold and 14 fold respectively). The significance of these findings in light of the recent literature is not clear and merits further investigation. It may either represent a mechanism of maintaining the cell against the effect of retinoic acid, the stimulus used for differentiation in these human EC experiments or support a role for CYP26A1 in the pluripotent phenotype as demonstrated by its over expression in pluripotent cells.

The second component of nullipotency involved a direct comparison of nulli- and pluripotent EC cells thus ultimately producing a transcriptome profile of nullipotency. This comparison yielded 1,575 genes upregulated in nullipotent cells with 1,031 genes downregulated when a FDR $< 0.05$ and fold change $> 2$ and $< 0.5$ were considered. Looking at the most highly up

and downregulated genes in nullipotent compared with pluripotent human EC cells, it included genes such as LCP1, HORMAD1, PIWIL2, MYH6 and CDH1. PIWIL2, one of the most highly downregulated genes and this fits with its reported property as a late marker of stem cells (Josephson et al., 2007). The Wnt signalling pathway was highly represented in the nullipotent data as might be expected as Wnt signalling has been described widely in hES and EC cells (Taipale et al., 2001).

In summary, analysis of nullipotency in human EC cells has produced a unique molecular signature of nullipotency with 1,575 gene upregulated in nullipotent cells compared with their pluripotent counterparts. CYP26A1 is an example of one differentially expressed gene found in both nullipotent differentiation and nullipotent versus pluripotent comparisons and its role needs to be clarified further to determine its precise role in the regulation of the cancer stem cell.

### 7.4 Comparison of mouse and human model

Much of the initial detailed molecular genetic analysis in research is performed on other more available and easier to manage mammalian systems than the human with the laboratory mouse model considered a superior model to most (Rosenthal et al., 2007). This work in turn highlights specific targets to further investigate in human cells with a view to unravelling the molecular detail of the human in the normal and diseased state. A large volume of knowledge in the stem cell field was initially derived from studies performed in mouse embryos and ES cells (Rao, 2004). Mouse models are commonly used as a representative mammalian model as used initially in this thesis and indeed as hES cells are not used for research purposes in this country. While many similarities are present thus allowing findings from mouse models to be potentially extrapolated to humans, a number of differences need to be considered.

#### 7.4.1 Differences and Similarities between the Mouse and Human Model

There are many fundamental similarities and differences between mouse and human ES and EC cells which are worthy of note and due consideration. The main differences between

mouse and human ES and EC cells include differences in their morphology, expression of cell surface markers (mES cells express SSEA1, hES cells express SSEA3 and 4) and their responses to LIF (Reubinoff et al., 2000). Mouse ES cells in vitro can use LIF to maintain pluripotency and thus preventing differentiation while hES cells do not. The reason underlying this difference is the LIF/Stat3 signalling pathway that operates in mES cells in response to the presence of LIF that is not present in hES cells (Brandenberger et al., 2004; Rao, 2004). Many of the components of LIF signalling were not shown to be detectable in hES cells except for Stat1 and Stat3, both of which were not, however, upregulated (Brandenberger et al., 2004).

Comparisons of the mouse and human cell properties are limited by the availability of cross-species arrays and the requirement for complex data analysis of vast quantities of data. All experiments in this thesis were performed using the AB array platform of which there is little if any written about in the published literature in the area of stem cell biology. Data analysis in the human and mouse models within this thesis provided the ortholog of the gene in other mammalian systems, namely mouse, human and rat. This allowed a cross-species comparison of the data.

***Similarities***

**Table 7.1: Similarities between mouse and human ES and EC cells.**

| **General similarities between mouse and human ES and EC cells** | |
|---|---|
| Expression of master regulators of self-renewal | Expression and requirement for Oct4, Sox2 and Nanog in the control of self-renewal |
| Signalling pathways | Wnt signalling a common pathway in maintaining undifferentiated state. |
| **Similarities found in this thesis** | |
| Genes differentially expressed upon differentiation in both human and mouse cells | The following genes; SEMA7A, LOXL1, EGR1, GSTO2, PRKAR1A; were found to be significantly differentially expressed in both mouse and human differentiation. |
| Genes differentially expressed in Nullipotency | Many genes were found to differentially expressed in Nullipotent cells in both the mouse and human model, e.g. Ctsh, Fxyd2, Crygd, Hpgd. |

Sato et al propose that Wnt signalling is a possible common pathway that maintains the undifferentiated state in hES and mES cells (Sato et al., 2004). They also showed that Wnt signalling positively regulated expression of Oct4 and Rex1 in both hES and mES cells. Nanog is also required. There are numerous genes that exert some form of control over pluripotency which may reflect its fundamental biological function within the cells and therefore was required to be evolutionally secured by multiple backup systems.

A comparison of differentially expressed genes upon differentiation in both human and mouse models uncovered 5 genes that were significant in both human and mouse, namely SEMA7A, LOXL1, EGF1, GSTO2 and PRKAR1A. Given the much larger number of differentially expressed genes determined in each individual comparison, the fact that only 5 common genes were found appears small. However, many of the genes in these comparisons are unknown and there were a significant number of genes with no corresponding ortholog. These 5 genes may, therefore, be worthy of further study.

Interestingly, a larger number of genes were common to both mouse and human when the comparison of genes differentially expressed between nulli and pluripotent EC cells was performed. This perhaps indicates that this comparison may be a vital source of genes important in pluripotency. Genes common to both models includes Ctsh, Fxyd1, Crygd and Hpgd with more genes highlighted in Table 6.17 and Table 6.18..

***Differences***

**Table 7.2: Differences between human and mouse models**

| | |
|---|---|
| **General human and mouse differences** | |
| ***LIF*** | hES cells do not respond or require LIF to prevent differentiation in vitro. mES cells can use LIF to prevent differentiation. |
| ***Expression of cell surface markers*** | hES cells express SSEA3 and 4. mES cells express SSEA1. |
| **Differences found in this thesis** | |
| ***Differentially expressed genes upon differentiation*** | The majority of differentially expressed genes upon differentiation of hEC cells were upregulated compared with the majority being downregulated in both mES and mEC cells. |
| ***PNCK*** | PNCK was the most upregulated gene in differentiated hEC cells. It was only marginally upregulated in differentiated mEC cells. |

PNCK, of which very little is published in the literature, was the highest upregulated gene found in differentiated Ntera2 cells. It was only marginally upregulated in differentiated SCC-PSA1 cells and it was not found to be differentially expressed in differentiated mES cells. Together, this data would support the findings of Perry Gardner et al, 2000, who concluded that PNCK may be expressed within a mammary epithelial cell type involved in differentiation as well as being involved in the transformation to the malignant phenotype (Perry Gardner et al., 2000).

### 7.4.2 Genes involved in Nullipotency

An identical comparison of nullipotent versus pluripotent embryonal carcinoma was performed in the mouse and human model with some similarities. Hydroxyprostaglandin dehydrogenase (HPGD), an enzyme involved in the degradation of prostaglandins, was found to be significantly upregulated in human nullipotent cells as well as mouse nullipotent cells.. In the human model, it was 15.5 fold upregulated while in the mouse model, it was one of the most highly upregulated genes in nullipotent cells at 43.0 fold. In the literature, HPGD has been described as downregulated in a metastasising oesophageal carcinoma cell line

(Kawamata et al., 2003) and downregulated in immortalised human urothelial cells (Chapman et al., 2008). Furthermore, its downregulation has also been reported in oral tongue squamous cell carcinoma (Ye et al., 2008), gastric cancer (Liu et al., 2008) and colorectal cancer ((Myung et al., 2006). In gastric cancer and lung cancer, it has been proposed that HPGD functions as a tumour suppressor gene (Tai et al., 2007). Its up regulation in this thesis in nullipotent embryonal carcinoma cells compared with its pluripotent counterpart runs contrary to the findings in the literature. Interestingly, in the mouse model, HPGD is upregulated in ES cells compared with pluripotent EC cells but downregulated in ES cells compared with nullipotent EC cells. It is clear that the function of HPGD in nullipotency i.e. in cells incapable of differentiation, and its role in stem cells and cancer stem cells needs to be further clarified as it may have some role to play in the maintenance of the self-renewing state and cell survival.

### 7.5 Microarray technology

Microarray technology is a relatively new and very powerful technology that is now widely used in biomedical research to analyse gene expression. DNA microarray analysis is unique in allowing large-scale gene expression profiling of thousands of genes in one experiment with a small amount of starting material thus making analysis of gene expression a much more economical and less laborious task. As well as all the advantages of using microarray technology, there are some associated limitations with doubts raised concerning repeatability, reproducibility and comparability of microarray technology. Since the introduction of DNA microarrays into the research environment, many different companies and laboratories have adopted numerous different protocols and array technologies. This has resulted in difficulties in array comparability with cross-platform comparisons of microarray data currently based on cross-referencing the annotation of each gene transcript represented on the arrays. This can often be an arduous task. Bioinformatics groups are investigating alternative approaches such as co-inertia analysis, a multivariate method that identifies trends or co-relationships in multiple datasets containing the same samples (Culhane et al., 2003). However, much debate still remains on the reproducibility of microarray experiments with blame placed on the technology used, biological and laboratory variation. Yauk et al, 2004 performed a comparative study on six different array platforms in an attempt to elucidate whether gene

expression profiles are more influenced by biology or the technological aspects of the experiment (Yauk et al., 2004). Their conclusion was that despite some technological differences that are important to consider and should be considered, the primary factor in determining variance was biological. Suarez-Farinas et al, 2005, performed a comparison of a number of microarray experiments on hESC to look at the question of comparison of independent microarray experiments (Suarez-Farinas et al., 2005). They concluded that the studies were comparable, compatible and repeatable but the publication of raw data rather than analysed data may be more useful for future comparison purposes. The MicroArray Quality Control (MAQC) project was launched by a group of US Food and Drug Administration (FDA) scientists to investigate the inter- and intraplatform reproducibility of gene expression experiments (Consortium, 2006). Overall, the microarray results between and within laboratories were comparable as well as being comparable across platforms thus supporting their continued use in gene expression profiling in basic and applied research. They also found that some platforms were more comparable to TaqMan® assays specifically Applied Biosystems and Agilent one-colour technologies. Dobbin et al, 2005, performed an inter and intra laboratory comparability study of cancer gene expression analysis to identify sources of variation in gene expression measurements from frozen tissues, cell line samples and purified RNA samples using four different laboratories (Dobbin et al., 2005). While they found the microarray data to be comparable between and within laboratories, they recommend that standardisation of the protocols for preparing and analysing samples is required to ensure comparability.

An additional limitation is the lack of standards for presenting and exchanging the vast array of data obtained from microarray studies. In 2001, a proposal was presented to remedy the situation and to ensure high quality microarray data called the Minimum Information About a Microarray Experiment (MIAME) describing the minimum information required to ensure that microarray data can be easily interpreted and that result derived from its analysis can be independently verified (Brazma et al., 2001). This proposal focuses on defining the content and structure of the required information rather than the technical format in capturing the data as well as ensuring public availability of all datasets. Meeting these minimum requirements for microarray experiments has now become almost compulsory for many journals prior to publication of any microarray experiment.

Affymetrix GeneChips are a commonly used microarray technology as well as Agilent technologies and other in house arrays. In this thesis, microarray technology from Applied Biosystems (AB) was used. In a direct comparison of the two platforms, the AB platform was shown to have a higher sensitivity, detecting four times as many gene changes and up to seven times as many when additional technical replicates were used (Ali-Seyed et al., 2006). Also, the AB microarray data correlated better with the qRT-PCR validation data ($R^2 = 0.71$) than the Affymetrix data ($R^2 = 0.47$).

Quantitative real-time RT-PCR (qRT-PCR) is often seen as the gold standard for gene expression measurement by biologists and is therefore often used to independently validate microarray results (Rajeevan et al., 2001b; Qin et al., 2006). It is a quick and quantitative method with small requirements for starting RNA. In one study by Rajeevan et al, 2001, microarray results were confirmed by qRT-PCR in 71% of cases with qRT-PCR usually finding greater expression differences than the microarray result (Rajeevan et al., 2001a). Microarray values and TaqMan® values correlate much better when there are high expression levels as demonstrated by Rajeevan et al, 2001 with the microarray results confirmed by qRT-PCR in 88% of cases when only those genes with high expression levels were considered. Most of the discrepancies that occur between TaqMan® based real-time PCR and microarrays are due the sensitivity limits between a PCR based and a hybridisation based approach (Wang et al., 2006b). Another reason may be that TaqMan® assays are targeting a subset of spliced transcripts that the microarrays measures.

### 7.6 Conclusions and Future Work

In conclusion, analysis of the cellular origin of cancer of different organs indicates that there is, in each instance, a determined stem cell that continues to divide into more stem cells and progeny that ultimately differentiate into the different cell types that make up the tumour thus maintaining the tumour (Chambers et al., 2004). Cancer stem cells have been identified in many different malignancies including breast cancer, melanoma, prostate cancer, brain tumours and most notably haematopoietic malignancies such as leukaemia where much of the initial work on the concept of the cancer stem cell was undertaken. Given that there is emerging evidence that stem cell biology may provide new insights into cancer biology, the

study of stem cells and their ability to self-renew and undergo differentiation may prove vitally important to understanding the regulation of cancer cell proliferation and progression. As well as prognostic implications, the cancer stem cell theory also has profound implications in the development of future cancer treatment strategies. The hope is that the specific targeting of cancer stem cells within a tumour as a treatment modality may provide more effective and successful cancer treatments than are currently available.

This thesis presents a model, namely teratoma tumourigenesis, in which to study cancer stem cells in solid tumours and their properties in comparison to their normal counterparts. Using high through-put microarray analysis, this thesis presents transcriptome profiles of 'benign' pluripotent stem cells and their 'malignant' counterpart as well as transcriptome profiles of cancer stem cell differentiation and self-renewal. Important biological processes and molecular functions involved in cancer stem cell proliferation and self-renewal are identified with a number of novel potential future targets for treatment identified. Array data was successfully validated using TaqMan® real-time quantitative RT-PCR. Data from the comparisons of mES cells and their malignant counterparts, embryonal carcinoma, also provides additional evidence to support the role of the cancer stem cell niche in the maintenance of the cancer stem cell.

**Table 7.3: Summary of data from Comparison of Differentiated and Undifferentiated cell groups.**

| Differentiation and Self-Renewal in ES and EC cells | | | | |
|---|---|---|---|---|
| | | ***Downregulated in differentiated cells*** | ***Upregulated in differentiated cells*** | ***Pathways and Molecular functions*** |
| ***Mouse ES cells*** | 592 gene signature (46 unique gene signature) | Dcn, Tnc, Ccl5, Oasl2, MP4, Pck1, Fbln2 | Mrgprh, Trim7, Lemd1, Bcl11b, Sema4d, Kcnj3, Actc1 | Cytokine signalling pathways, Extracellular matrix proteins, Signalling molecules |
| ***Mouse pluripotent EC cells*** | 224 gene signature (20 unique gene signature)) | Ccl5, Sdpr, Clec2d, Gria1, Ogn, Oasl2, Dcn, Tgfb3, Tnc, Tnfrsf9, Adamts5, Fbln5, Fbln2 | Olfr1450, Fgf5, Igfbp5, Afp, Mal2, Tmprss2, Dscam | Angiogenesis, Cytokines, Signalling molecules, extracellular matrix |
| ***Mouse nullipotent EC cells*** | No genes significantly differentially expressed | | | |
| ***Human pluripotent EC cells*** | 249 gene signature | TMC2, AKAP5, OPTC, TDGF1, SCAP2 | PNCK, POLI, TSPAN18, HEY2, MAPK13 | Angiogenesis, Signalling molecules |
| ***Human nullipotent EC cells*** | 12 gene signature | FOXD1, LHX6, GPR143 | CYP26A1, ZNF503, HOXB5, CD22, NMU | |

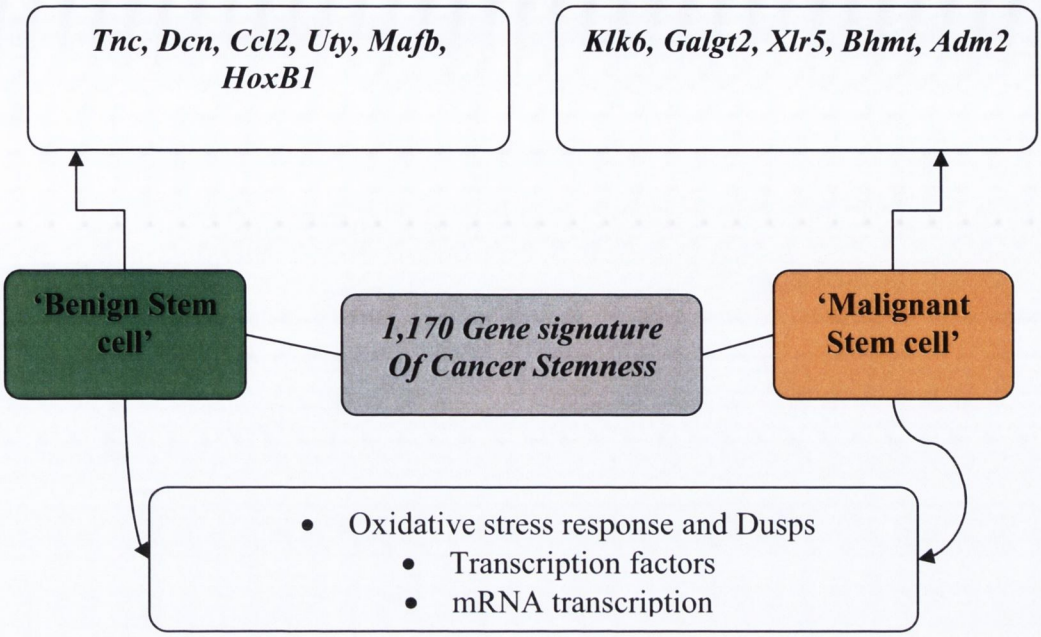



**Figure 7.2: Summary of comparison of 'benign' stem cell and 'malignant' or cancer stem cell, from the mouse model.**

Many of the differences between signalling pathways in normal and cancer stem cells have not yet been definitively elucidated (Zhang et al., 2006). Through the transcriptome profiling of both normal and cancer stem cells using teratoma tumourigenesis as the model system, new light has been shed on the similarities and differences between the two cell types. The data obtained in this thesis provides further evidence to support the cancer stem cell model and the important role of the cancer stem cell niche as well as identifying new targets that may potentially revolutionise the treatment of cancer.

Future work would involve further molecular dissection of targets uncovered by this transcriptional profiling work. This would enable the determination of the precise roles of the isolated targets with a view to developing future therapies that would selectively target and treat cancer stem cells. Gene silencing experiments would be of use in the characterisation of these potential targets to determine the effects of interruption or suppression of the gene at the transcriptional or translational level.

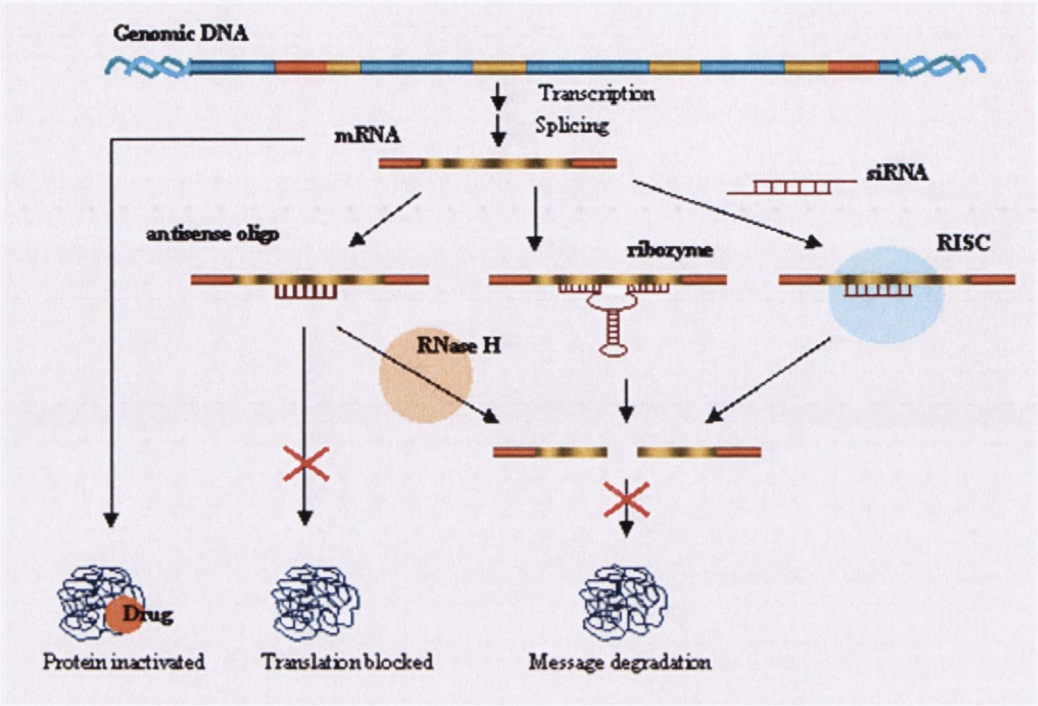



**Figure 7.3: Methods of gene silencing. (Taken from http://www.ncbi.nlm.nih.gov/projects/genome/probe/doc/ApplSilencing.shtml).**

A popular method of gene silencing is the use of antisense and RNA interference (RNAi) referred to as gene knockdown technologies. Here, the transcription of the gene is unaffected, however gene expression, i.e. protein synthesis, is lost due to mRNA molecules becoming unstable or inaccessible. A future aim following the production of a transcriptional profile of cancer stemness and elucidation of genes vital to cancer stem cells is to remove the cancer stem cell properties from the cancer cells via knockdown of key cancer stemness genes and miRNAs. Following performing this work in the current teratocarcinoma model, similar experiments would be performed with other tumour models within the group's research remit, namely cervical, ovarian and thyroid cancers.

A relatively recent discovery has been the regulation of mRNA molecules by small RNA species referred to small interfering RNA (siRNA) and microRNA (miRNA) which can bind to their target transcripts and repress their translation (Krutzfeldt et al., 2006; Rana, 2007). MiRNA profiles have been used to classify human cancers with differential miRNA

expression demonstrated in benign and malignant tissues as well as differentiated and undifferentiated tissues (Calin et al., 2006).

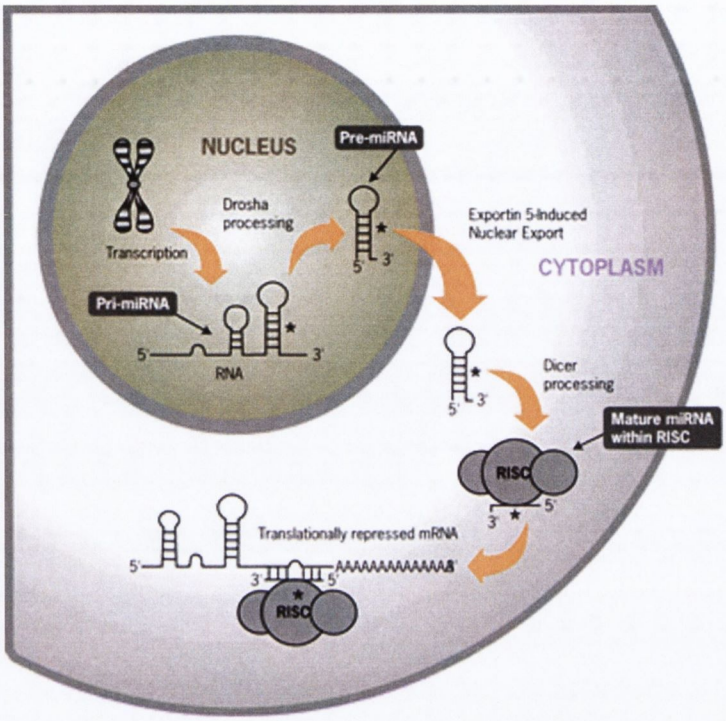



**Figure 7.4: MicroRNAs – Function and location within the cell. (Taken from Ambion®, www.ambion.com).**

These molecules may have a significant impact on the regulation of mRNA expression and therefore in the control of stem cell differentiation and self-renewal. Future work, which has in fact already begun, involves obtaining miRNA profiles of all samples used in the current teratocarcinoma model to try and gain a list of differentially expressed miRNA molecules and correlate these to the corresponding mRNA expression profiles already obtained.

# REFERENCES

Abate-Shen, C (2006). A New Generation of Mouse Models of Cancer for Translational Research. Clinical Cancer Research; **12**, 5274 - 5275.

Abe, K, Niwa, H, Iwase, K, Takiguchi, M, Mori, M, Abe, S-I, Abe, K, and Yamamura, K-I (1996). Endoderm-specific Gene Expression in Embryonic Stem Cells Differentiated to Embryoid Bodies. Experimental Cell Research; **229**, 27-34.

Abeyta, MJ, Clark, AT, Rodriguez, RT, Bodnar, MS, Pera, RAR, and Firpo, MT (2004). Unique Gene Expression Signatures of Independently-Derived Human Embryonic Stem Cell Lines. Human Molecular Genetics; **13**, 601-608.

Adams, GB, Martin, RP, Alley, IR, Chabner, KT, Cohen, KS, Calvi, LM, Kronenberg, HM, and Scadden, DT (2007). Therapeutic Targeting of a Stem Cell Niche. Nature Biotechnology; **25**, 238-243.

Ahn, J-I, Lee, K-H, Shin, D-M, Shim, J-W, Lee, J-S, Chang, SY, Lee, Y-S, Brownstein, MJ, Lee, S-H, and Lee, Y-S (2004). Comprehensive Transcriptome Analysis of Differentiation of Embryonic Stem Cells Into Midbrain and Hindbrain Neurons. Developmental Biology; **265**, 491-501.

Al-Hajj, M, Wicha, MS, Benito-Hernandez, A, Morrison, SJ, and Clarke, MF (2003). Prospective Identification of Tumorigenic Breast Cancer Cells. Proc Natl Acad Sci U S A; **100**, 3983-8.

Albig, AR, Neil, JR, and Schiemann, WP (2006). Fibulins 3 and 5 Antagonize Tumor Angiogenesis *In vivo*. Cancer Research; **66**, 2621-2629.

Ali-Seyed, M, Laycock, N, Karanam, S, Xiao, W, Blair, ET, and Moreno, CS (2006). Cross-Platform Expression Profiling Demonstrates that SV40 Small Tumor Antigen Activates Notch, Hedgehog, and Wnt signaling in Human Cells. BMC Cancer; **6**.

Amoureux, M-C, Cunningham, BA, Edelman, GM, and Crossin, KL (2000). N-CAM Binding Inhibits the Proliferation of Hippocampal Progenitor Cells and Promotes Their Differentiation to a Neuronal Phenotype. Journal of Neuroscience; **20**, 3631-3640.

Andrews, PW (1998). Teratocarcinomas and Human Embryology: Pluripotent Human EC Cell Lines. APMIS; **106**, 158-168.

Andrews, PW (2002). From Teratocarcinomas to Embryonic Stem Cells. Philos Trans R Soc Lond B Biol Sci; **357**, 405-17.

Andrews, PW, Goodfellow, PN, Shevinsky, LH, Bronson, DL, and Knowles, BB (1982). Cell-surface Antigens of a Clonal Human Embryonal Carcinoma Cell Line: Morphological and Antigenic Differentiation in Culture. Int J Cancer; **29**, 523-531.

Andrews, PW, Matin, MM, Bahrami, AR, Damjanov, I, Gokhale, P, and Draper, JS (2005). Embryonic Stem (ES) Cells and Embryonal Carcinoma (EC) Cells: Opposite Sides of the Same Coin. Biochemical Society Transactions; **33**, 1526-1529.

Astigiano, S, Damonte, P, Fossati, S, Boni, L, and Barbieri, O (2005). Fate of Embryonal Carcinoma Cells Injected Into Postimplantation Mouse Embryos. Differentiation; **73**, 484-490.

Ben-Porath, I, Thomson, MW, Carey, VJ, Ge, R, Bell, GW, Regev, A, and Weinberg, RA (2008). An Embryonic Stem Cell-Like Gene Expression Signature in Poorly Differentiated Aggressive Human Tumours. Nature Genetics; **40**, 499-507.

Benjamini, Y, Drai, D, Elmer, G, Kafkafi, N, and Golani, I (2001). Controlling the false discovery rate in behavior genetics research. Behav Brain Res; **125**, 279-84.

Bhattacharya, B, Cai, J, Luo, Y, Miura, T, Mejido, J, Brimble, SN, Zeng, X, Schulz, TC, Rao, MS, and Puri, RK (2005). Comparison of the Gene Expression Profile of Undifferentiated Human Embryonic Stem Cell Lines and Differentiating Embryoid Bodies. BMC Developmental Biology; **5**.

Bhattacharya, B, Miura, T, Brandenberger, R, Mejido, J, Luo, Y, Yang, AX, Joshi, BH, Ginis, I, Thies, RS, Amit, M, Lyons, I, Condie, BG, Itskovitz-Eldor, J, Rao, MS, and Puri, RK (2004). Gene Expression in Human Embryonic Stem Cell Lines: Unique Molecular Signature. Blood; **103**, 2956-2964.

Bjerkvig, R, Tysnes, BB, Aboody, KS, Najbauer, J, and Terzis, AJA (2005). The Origin of the Cancer Stem Cell: Current Controversies and New Insights. Nature Reviews Cancer; **5**, 899-904.

Bonnet, D, and Dick, JE (1997). Human Acute Myeloid Leukemia is Organised as a Hierarchy that Originates from a Primitive Haematopoietic Cell. Nature Medicine; **3**, 730-737.

Borczuk, AC, Papanikolaou, N, Toonkel, RL, Sole, M, Gorenstein, LA, Ginsburg, ME, Sonett, JR, Friedman, RA, and Powell, CA (2008). Lung Adenocarcinoma Invasion In TGFBRII-Deficient Cells is Mediated by CCL5/RANTES. Oncogene; **27**, 557-564.

Brandenberger, R, Khrebtukova, I, Thies, RS, Miura, T, Jingli, C, Puri, RK, Vasicek, T, Lebkowski, J, and Rao, M (2004a). MPSS Profiling of Human Embryonic Stem Cells. BMC Developmental Biology; **4**, 10.

Brandenberger, R, Wei, H, Zhang, S, Lei, S, Murage, J, Fisk, GJ, Li, Y, Xu, C, Fang, R, Guegler, K, Rao, MS, Mandalam, R, Lebkowski, J, and Stanton, LW (2004b). Transcriptome Characterization Elucidates Signaling Networks that Control Human ES Cell Growth and Differentiation. Nature Biotechnology; **22**, 707-716.

Brazma, A, Hingamp, P, Quackenbush, J, Sherlock, G, Spellman, P, Stoeckert, C, Aach, J, Ansorge, W, Ball, CA, Causton, HC, Gaasterland, T, Glenisson, P, Holstege, FCP, Kim, IF, Markowitz, V, Matese, JC, Parkinson, H, Robinson, A, Sarkans, U, Schulze-Kremer, S, Stewart, J, Taylor, R, Vilo, J, and Vingron, M (2001). Minimum Information About A Microarray Experiment (MIAME) - Toward Standards For Microarray Data. Nature Genetics; **29**, 365-371.

Calin, GA, and Croce, CM (2006). MicroRNA Signatures in Human Cancers. Nature Reviews Cancer; **6**, 857-866.

Capkovic, KL, Stevenson, S, Johnson, MC, Thelen, JJ, and Cornelison, DDW (2008). Neural Cell Adhesion Molecule (NCAM) Marks Adult Myogenic Cells Committed to Differentiation. Experimental Cell Research; **314**, 1553-1565.

Carpenter, MK, Inokuma, MS, Denham, J, Mujtaba, T, Chiu, C-P, and Rao, MS (2001). Enrichment of Neurons and Neural Precursors from Human Embryonic Stem Cells. Experimental Neurology; **172**, 383-397.

Chambers, I, and Smith, A (2004). Self-renewal of Teratocarcinoma and Embryonic Stem Cells. Oncogene; **23**, 7150-7160.

Chang, C-L, Hong, E, Lao-Sirieix, P, and Fitzgerald, RC (2008). A Novel Role for the Retinoic Acid-Catabolising Enzyme CYP26A1 in Barrett's Associated Adenocarcinoma. Oncogene; **27**, 2951-2960.

Chapman, EJ, Kelly, G, and Knowles, MA (2008). Genes Involved in Differentiation, Stem Cell Renewal, and Tumourigenesis Are Modulated in Telomerase-Immortalised Human Urothelial Cells. Molecular Cancer Research; **6**, 1154-1168.

Chen, S-H, Pham-Nguyen, KB, Martinet, O, Huang, Y, Yang, W, Thung, SN, Chen, L, Mittler, RS, and Woo, SLC (2000). Rejection of Disseminated Metastases of Colon Carcinoma by Synergism of IL-12 Gene Therapy and 4-1BB Costimulation. Molecular Therapy; **2**, 39-46.

Chinzei, R, Tanaka, Y, Shimizu-Saito, K, Hara, Y, Kakinuma, S, Watanabe, M, Teramoto, K, Arii, S, Takase, K, Sato, C, Terada, N, and Teraoka, H (2002). Embryoid-Body Cells Dervied From a Mouse Embryonic Stem Cell Line Show Differentiation Into Functional Hepatocytes. Hepatology; **36**, 22-29.

Choi, D, Lee, H-J, Jee, S, Jin, S, Koo, SK, Paik, SS, Jung, SC, Hwang, S-Y, Lee, KS, and Oh, B (2005). In Vitro Differentiation of Mouse Embryonic Stem Cells: Enrichment of Endodermal Cells in the Embryoid Body. Stem Cells; **23**, 817-827.

Christie, GR, Williams, DJ, Macisaac, F, Dickinson, RJ, Rosewell, I, and Keyse, SM (2005). The Dual-Specificity Protein Phosphatase DUSP9/MKP-4 is Essential for Placental Function But is Not Required for Normal Embryonic Development. Mol Cell Biol; **25**, 8323-33.

Cinelli, P, Casanova, EA, Uhlig, S, Lochmatter, P, Matsuda, T, Yokota, T, Rulicke, T, Ledermann, B, and Burki, K (2008). Expression Profiling in Transgenic FVB/N Embryonic Stem Cells Overexpressing STAT3. BMC Developmental Biology; **8**, 57.

Clarke, MF (2005). Self-Renewal and Solid-Tumor Stem Cells. Biology of Blood and Marrow Transplantation; **11**, 14-16.

Collins, AT, Berry, PA, Hyde, C, Stower, MJ, and Maitland, NJ (2005). Prospective Identification of Tumourigenic Prostate Cancer Stem Cells. Cancer Research; **65**, 10946-10951.

Collins, AT, and Maitland, NJ (2006). Prostate Cancer Stem Cells. European Journal of Cancer; **42**, 1213-1218.

Consortium, M (2006). The MicroArray Quality Control (MAQC) Project Shows Inter- and Intraplatform Reproducibility of Gene Expression Measurements. Nature Biotechnology; **24**, 1151-1161.

Cortes, F, Debacker, C, Peault, B, and Labastie, M-C (1999). Differential Expression of KDR/VEGFR-2 and CD34 During Mesoderm Development of the Early Human Embryo. Mechanisms of Development; **83**, 161-164.

Culhane, AC, Perriere, G, and Higgins, DG (2003). Cross-platform Comparison and Visualisation of Gene Expression Data Using Co-Inertia Analysis. BMC Bioinformatics; **4**, 59.

Darr, H, Mayshar, Y, and Benvenisty, N (2006). Overexpression of NANOG in Human ES cells Enables Feeder-Free Growth While Inducing Primitive Ectoderm Features. Development; **133**, 1193-1201.

Dean, M, Fojo, T, and Bates, S (2005). Tumour Stem Cells and Drug Resistance. Nature Reviews Cancer; **5**, 275-284.

Deb, TB, Coticchia, C, M., Barndt, R, Zuo, H, Dickson, RB, and Johnson, MD (2008). Pregnancy-upregulated Nonubiquitous Calmodulin Kinase Induces Ligand-independent EGFR Degradation. Am J Physiol Cell Physiol; **295**, 365-377.

Degenhardt, K, and White, E (2006). A Mouse Model System to Genetically Dissect the Molecular Mechanisms Regulating Tumorigenesis. Clinical Cancer Research; **12**, 5298-5304.

Dobbin, KK, Beer, DG, Meyerson, M, Yeatman, TJ, Gerald, WL, Jacobson, JW, Conley, B, Buetow, KH, Heiskanen, M, Simon, RM, Minna, JD, Girard, L, Misek, DE, Taylor, JMG, Hanash, S, Naoki, K, Hayes, DN, Ladd-Acosta, C, Enkemann, SA, Viale, A, and Giordano, TJ (2005). Interlaboratory Comparability Study of Cancer Gene Expression Analysis Using Oligonucleotide Microarrays. Clinical Cancer Research; **11**, 565-572.

Doss, MX, Koehler, CI, Gissel, C, Hescheler, J, and Sachinidis, A (2004). Embryonic Stem Cells: A Promising Tool for Cell Replacement Therapy. Journal of Cellular Molecular Medicine; **8**, 465-473.

Douglass, EC, Green, AA, Hayes, FA, Etcubanas, E, Horowitz, M, and Wilimas, JA (1985). Chromosome 1 Abnormalities: A Common Feature of Paediatric Solid Tumours. J Natl Cancer Inst; **75**, 51-54.

Downing, GJ, and Battey, JFJ (2004). Technical Assessment of the First 20 Years of Research Using Mouse Embryonic Stem Cell Lines. Stem Cells; **22**, 1168-1180.

Duran, C, Talley, PJ, Walsh, J, Pigott, C, Morton, IE, and Andrews, PW (2001). Hybrids of Pluripotent and Nullipotent Human Embryonal Carcinoma Cells: Partial Retention of a Pluripotent Phenotype. Int J Cancer; **93**, 324-332.

Eichmann, A, Corbel, C, Nataf, V, Vaigot, P, Breant, C, and Le Douarin, NM (1997). Ligand-Dependent Development of the Endothelial and Haemopoietic Lineages from Embryonic Mesodermal Cells Expressing Vascular Endothelial Growth Factor Receptor 2. Proc Natl Acad Sci USA; **94**, 5141-5146.

Epping, M, and Bernards, R (2006). A Causal Role for the Human Tumour Antigen Preferentially Expressed Antigen of Melanoma in Cancer. Cancer Research; **66**, 10639-10642.

Eskildsen, S, Justesen, J, Schierup, MH, and Hartmann, R (2003). Characterisation of the 2'-5'- Oligoadenylate Synthetase Ubiquitin-Like Family. Nucleic Acids Research; **31**, 3166-3173.

Fleige, S, and Pfaffl, MW (2006). RNA Integrity and the Effect on the Real-time qRT-PCR Performance. Molecular Aspects of Medicine; **27**, 126-139.

Foell, JL, Diez-Mendiondo, BI, Diez, OH, Holzer, U, Ruck, P, Bapat, AS, Hoffmann, MK, Mittler, RS, and Dannecker, GE (2004). Engagement of the CD137 (4-1BB) Costimulatory Molecule Inhibits and Reverses the Autoimmune Process in Collagen-Induced Arthritis and Established Lasting Disease Resistance. Immunology; **113**, 89-98.

Fomchenko, EI, and Holland, EC (2006). Mouse Models of Brain Tumours and Their Applications in Preclinical Trials. Clinical Cancer Research; **12**, 5288-5297.

Frese, KK, and Tuveson, DA (2007). Maximizing Mouse Cancer Models. Nature Reviews Cancer; **7**, 645-658.

Gardner, LA, Tavalin, SJ, Goehring, AS, Scott, JD, and Bahouth, SW (2006). AKAP79-mediated Targeting of the Cyclic AMP-dependent Protein Kinase to the B1-

Adrenergic Receptor Promotes Recycling and Functional Resensitisation of the Receptor. Journal of Biological Chemistry; **281**, 33537-33553.

Gendron, C, Kashlwagl, M, Llm, NH, Enghlld, JJ, Thogersen, IB, Hughes, C, Caterson, B, and Nagase, H (2007). Proteolytic Activities of Human ADAMTS-5 Comparative Studies with ADAMTS-4. Journal of Biological Chemistry; **282**, 18294-18306.

Gerard, C, and Rollins, BJ (2001). Chemokines and Disease. Nature Immunology; **2**, 108-115.

Gibbs, CP, Kukekov, VG, Reith, JD, Tchigrinova, O, Suslov, ON, Scott, EW, Ghivizzani, SC, Ignatova, TN, and Steindler, DA (2005). Stem-Like Cells in Bone Sarcomas: Implications for Tumorigenesis. Neoplasia; **7**, 967-976.

Ginis, I, Luo, Y, Miura, T, Thies, S, Brandenberger, R, Gerecht-Nir, S, Amit, M, Hoke, A, Carpenter, MK, Itskovitz-Eldor, J, and Rao, MS (2004). Differences Between Human and Mouse Embryonic Stem Cells. Developmental Biology; **269**, 360-380.

Graham, ME, Kilby, DM, Firth, SM, Robinson, PJ, and Baxter, RC (2007). The In Vivo Phosphorylation and Glycosylation of Human Insulin-like Growth Factor-binding Protein-5. Molecular and Cellular Proteomics; **6**, 1392-1405.

Green, JE, and Hudson, T (2005). The Promise of Genetically Engineered Mice for Cancer Prevention. Nature Reviews Cancer; **5**, 184-198.

Gudjonsson, T, and Magnusson, MK (2005). Stem Cell Biology and The Cellular Pathways of Carcinogenesis. APMIS; **113**, 922-9.

Hailesellasse Sene, KH, Porter, CJ, Palidwor, G, Perez-Iratxeta, C, Muro, EM, Campbell, PA, Rudnicki, MA, and Andrade-Navarro, MA (2007). Gene Function in Early Mouse Embryonic Stem Cell Differentiation. BMC Genomics; **8**, 53.

Hamer, G, Gell, K, Kouznetsova, A, Novak, I, Benavente, R, and Hoog, C (2006). Characterisation of a Novel Meiosis-Specific Protein Within the Central Element of the Synaptonemal Complex. Journal of Cell Science; **119**, 4025-4032.

Hanahan, D, and Weinberg, RA (2000). The Hallmarks of Cancer. Cell; **100**, 57-70.

Harada, T, Chelala, C, Bhakta, V, Chaplin, T, Caulee, K, Baril, P, Young, BD, and Lemoine, NR (2008). Genome-wide DNA Copy Number Analysis in Pancreatic

Cancer Using High-Density Single Nucleotide Polymorphism Arrays. Oncogene; **27**, 1951-1960.

Held-Feindt, J, Paredes, EB, Blomer, U, Seidenbecher, C, Stark, AM, Mehdorn, HM, and Mentlein, R (2005). Matrix-Degrading Proteases ADMATS4 and ADAMTS5 (disintegrins and metalloproteinases with thrombospondin motifs 4 and 5) are Expressed in Human Glioblastomas. International Journal of Cancer; **118**, 55-61.

Heo, J, Lee, J-S, Chu, I-S, Takahama, Y, and Thorgeirsson, SS (2005). Spontaneous Differentiation of Mouse Embryonic Stem Cells In Vitro: Characterisation by Global Gene Expression Profiles. Biochemical and Biophysical Research Communications; **332**, 1061-1069.

Hirashima, M, Ogawa, M, Nishikawa, S, Matsumura, K, Kawasaki, K, Shibuya, M, and Nishikawa, S-I (2003). A Chemically Defined Culture of VEGFR2+ Cells Derived From Embryonic Stem Cells Reveals The Role of VEGFR1 in Tuning the Threshold for VEGF in Developing Endothelial Cells. Blood; **101**, 2261-2267.

Hiratsuka, S, Nakao, K, Nakamura, K, Katsuki, M, Maru, Y, and Shibuya, M (2005). Membrane Fixation of Vascular Endothelial Groth Factor Receptor 1 Ligand-Binding Domain is Important for Vasculogenesis and Angiogenesis in Mice. Molecular and Cellular Biology; **25**, 346-354.

Hoornaert, I, Marynen, P, Goris, J, Sciot, R, and Baens, M (2003). MAPK Phosphatase DUSP16/MKP-7, a Candidate Tumour Suppressor for Chromosome Region 12p12-13, Reduces BCR-ABL-Induced Transformation. Oncogene; **22**, 7728-36.

Huff, CA, Matsui, WH, Smith, BD, and Jones, RJ (2006). Strategies to Eliminate Cancer Stem Cells: Clinical Implications. European Journal of Cancer; **42**, 1293-1297.

Iozzo, RV (1998). Matrix Proteoglycans: From Molecular Design to Cellular Function. Annu Rev Biochem; **67**, 609-52.

Ivanova, NB, Dimos, JT, Schaniel, C, Hackney, JA, Moore, KA, and Lemischka, IR (2002). A Stem Cell Molecular Signature. Science; **298**, 601-604.

Izumi, N, Era, T, Akimaru, H, Yasunaga, M, and Nishikawa, S-I (2007). Dissecting the Molecular Hierarchy for Mesendoderm Differentiation Through a Combination of Embryonic Stem Cell Culture and RNA Interference. Stem Cells; **25**, 1664-1674.

Johnson, SK, Dennis, RA, Barone, GW, Lamps, LW, and Haun, RS (2006). Differential Expression of Insulin-like Growth Factor Binding Protein-5 in Pancreatic Adenocarcinomas: Identification using DNA Microarray. Molecular Carcinogenesis; **45**, 814-827.

Jones, DL, and Wagers, AJ (2008). No Place Like Home: Anatomy and Function of the Stem Cell Niche. Nature Reviews Molecular Cell Biology; **9**, 11-21.

Josephson, R, Ording, CJ, Liu, Y, Shin, S, Lakshmipathy, U, Toumadje, A, Love, B, Chesnut, JD, Andrews, PW, Rao, M, and Auerbach, JM (2007). Qualification of Embryonal Carcinoma 2102Ep As a Reference for Human Embryonic Stem Cell Research. Stem Cells; **25**, 437-446.

Kaji, K, Caballero, IM, MacLeod, R, Nichols, J, Wilson, VA, and Hendrich, B (2006). The NuRD Component Mbd3 is Required for Pluripotency of Embryonic Stem Cells. Nature Cell Biology; **8**, 285-292.

Kakinuma, T, and Hwang, ST (2006). Chemokines, Chemokine Receptors, and Cancer Metastasis. Journal of Leukocyte Biology; **79**, 639-651.

Kanehisa, M, and Goto, S (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research; **28**, 27-30.

Kawamata, H, Furihata, T, Omotehara, F, Sakai, T, Horiuchi, H, Shinagawa, Y, Imura, J, Ohkura, Y, Tachibana, M, Kubota, K, Terano, A, and Fujimori, T (2003). Identification of Genes Differentiallly expressed ikn a Newly Isolated Human Metastasising Oesophageal Cancer Cell Line, T.Tn-AT1, by cDNA Microarray. Cancer Science; **94**, 699-706.

Keresztes, G, Mutai, H, and Heller, S (2003). TMC and EVER Genes Belong to a Larger Novel Family, the TMC Gene Family Encoding Transmembrane Proteins. BMC Genomics; **4**, 24.

Kevorkian, L, Young, DA, Darrah, C, Donell, ST, Shepstone, L, Porter, S, Brockbank, SMV, Edwards, DR, Parker, AE, and Clark, IM (2004). Expression Profiling of Metalloproteinases and Their Inhibitors in Cartilage. Arthritis and Rheumatism; **50**, 131-141.

Kim, JA, Averbook, BJ, Chambers, K, Rothchild, K, Kjaergaard, J, Papay, R, and Shu, S (2001). Divergent Effects of 4-1BB Antibodies on Antitumour Immunity and on Tumour-Reactive T-Cell Generation. Cancer Research; **61**, 2031-2037.

Knudsen, S (2004): *Guide to analysis of DNA microarray data.* Wiley-Liss. Hoboken, N.J.

Koestenbauer, S, Zech, NH, Juch, H, Vanderzwalmen, P, Schoonjans, L, and Dohr, G (2006). Embryonic Stem Cells: Similarities and Differences Between Human and Murine embryonic Stem Cells. American Journal of Reproductive Immunology; **55**, 169-180.

Koninger, J, Giese, NA, di Mola, FF, Berberat, P, Giese, T, Esposito, I, Bachem, MG, Buchler, MW, and Friess, H (2004). Overexpressed Decorin in Pancreatic Cancer: Potential Tumour Growth Inhibition and Attenuation of Chemotherapeutic Action. Clinical Cancer Research; **10**, 4776-4783.

Krawetz, R, and Kelly, GM (2007). Wnt6 Induces the Specification and Epithelialisation of F9 Embryonal Carcinoma Cells to Primitive Endoderm. Cellular Signalling; **20**, 506-517.

Krutzfeldt, J, Poy, MN, and Stoffel, M (2006). Strategies to Determine the Biological Function of MicroRNAs. Nature Genetics; **38**, S14-S19.

Kuai, XL, Cong, XQ, Li, XL, and Xiao, SD (2003). Generation of Hepatocytes from Cultured Mouse Embryonic Stem Cells. Liver Transplantation; **9**, 1094-1099.

Kubista, M, Andrade, JM, Bengtsson, M, Forootan, A, Jonak, J, Lind, K, Sindelka, R, Sjoback, R, Sjogreen, B, Strombom, L, Stahlberg, A, and Zoric, N (2006). The Real-Time Polymerase Chain Reaction. Molecular Aspects of Medicine; **27**, 95-125.

Kurima, K, Yang, Y, Sorber, K, and Griffith, AJ (2003). Characterisation of the Transmembrane Channel-like (TMC) Gene Family: Functional Clues From Hearing Loss and Epidermodysplasia Verruciformis. Genome; **82**, 300-308.

Kuroda, T, Tada, M, Kubota, H, Kimura, H, Hatano, S-y, Suemori, H, Nakatsuji, N, and Tada, T (2005). Octamer and Sox Elements Are Required for Transcriptional cis Regulation of Nanog Gene Expression. Molecular and Cellular Biology; **25**, 2475-2485.

Langton, S, and Gudas, LJ (2008). CYP26A1 Knockout Embryonic Stem Cells Exhibit Reduced Differentiation and Growth Arrest in Response to Retinoic Acid. Developmental Biology; **315**, 331-354.

Lapidot, T, Sirard, C, Vormoor, J, Murdoch, B, Hoang, T, Caceres-Cortes, J, Minden, M, Paterson, B, Caligiuri, M, and Dick, JE (1994). A cell Initiating Human Acute Myeloid Leukaemia After Transplantation Into SCID Mice. Nature; **367**, 645-648.

Lee, S-W, Vella, AT, Kwon, BS, and Croft, M (2005). Enhanced CD4 T Cell Responsiveness in the Absence of 4-1BB. Journal of Immunology; **174**, 6803-6808.

Li, W, and Guan, K-L (2004). The Down Syndrome Cell Adhesion Molecule (DSCAM) Interacts with and Activates Pak. Journal of Biological Chemistry; **279**, 32824-32836.

Liu, Y, Shin, S, Zeng, X, Zhan, M, Gonzalez, R, Mueller, F-J, Schwartz, CM, Xue, H, Li, H, Baker, SC, Chudin, E, Barker, DL, McDaniel, TK, Oeser, S, Loring, JF, Mattson, MP, and Rao, MS (2006). 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 Developmental Biology; **6**, 20.

Liu, Z, Wang, X, Lu, Y, Han, S, Zhang, F, Zhai, H, Lei, T, Liang, J, Wang, J, Wu, K, and Fan, D (2008). Expression of 15-PGDH is Downregulated by COX-2 in Gastric Cancer. Carcinogenesis; **29**, 1219-1227.

Livak, KJ, and Schmittgen, TD (2001). Analysis of Relative Gene Expression Data using Real-Time Quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods; **25**, 402-408.

Looijenga, LH, Zafarana, G, Grygalewicz, B, Summersgill, B, Debiec-Rychter, M, Veltman, J, Shchoenmakers, EF, Rodriguez, S, Jafer, O, Clark, J, van Kessel, AG, Shipley, J, van Gurp, RJ, Gillis, AJ, and Oosterhuis, JW (2003). Role of Gain of 12p in Germ Cell Tumour development. APMIS; **111**, 161-171.

Lyon, MA, Ducruet, AP, Wipf, P, and Lazo, JS (2002). Dual-Specificity Phosphatases as Targets for Antineoplastic Agents. Nature Reviews Drug Discovery; **1**, 961-976.

Martin, GR, and Evans, MJ (1975). Differentiation of Clonal Lines of Teratocarcinoma Cells: Formation of Embryoid Bodies *In Vitro*. Proc Nat Acad Sci USA; **72**, 1441-1445.

Martinet, O, Divino, CM, Zang, Y, Gan, Y, Mandeli, J, Thung, S, Pan, P-Y, and Chen, S-H (2002). T Cell Activation with Systemic Agonistic Antibody Versus Local 4-1BB Ligand Gene Delivery Combined With Interleukin-12 Eradicate Liver Metastases of Breast. Gene Therapy; **9**, 786-792.

Maser, RS, Choudhury, B, Campbell, PJ, Feng, B, Wong, K-K, Protopopov, A, O'Neil, J, Gutierrez, A, Ivanova, E, Perna, I, Lin, E, Mani, V, Jiang, S, McNamara, K, Zaghlul, S, Edkins, S, Stevens, C, Brennan, C, Martin, ES, Wiedemeyer, R, Kabbarah, O, Nogueira, C, Histen, G, Aster, JC, Mansour, M, Duke, V, Foroni, L, Fielding, AK, Goldstone, AH, Rowe, JM, Wang, YA, Look, AT, Stratton, MR, Chin, L, Futreal, PA, and DePinho, RA (2007). Chromosomally Unstable Mouse Tumours Have Genomic Alaterations Similar To Diverse Human Cancers. Nature; **447**, 966-971.

Masui, S, Nakatake, Y, Toyooka, Y, Shimosato, D, Yagi, R, Takahashi, K, Okochi, H, Okuda, A, Matoba, R, Sharov, AA, Ko, MSH, and Niwa, H (2007). Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. Nature Cell Biology; **9**, 625-635.

Miyatake, T, Ueda, Y, Nakashima, R, Yoshino, K, Kimura, T, Murata, T, Nomura, T, Fujita, M, Buzard, GS, and Enomoto, T (2007). Down-regulation of Insulin-like Growth Factor Binding Protein-5 (IGFBP-5): Novel Marker for Cervical Carcinogenesis. International Journal of Cancer; **120**, 2068-2077.

Myung, S-J, Rerko, RM, Yan, M, Platzer, P, Guda, K, Dotson, A, Lawrence, E, Dannenberg, AJ, Lovgren, AK, Luo, G, Pretlow, TP, Newman, RA, Willis, J, Dawson, D, and Markowitz, SD (2006). 15-Hydroxyprostaglandin Dehydrogenase is an In Vivo Suppressor of Colon Tumourigenesis. PNAS; **103**, 12098-12102.

Nagase, H, and Kashiwagi, M (2003). Aggrecanases and Cartilage Matrix Degradation. Arthritis Research and Therapy; **5**, 94-103.

Nakada, M, Miyamori, H, Kita, D, Takahashi, T, Yamashita, J, Sato, H, Miura, R, Yamaguchi, Y, and Okada, Y (2005). Human glioblastomas overexpress ADAMTS-5 that degrades brevican. Acta Neuropatholgica; **110**, 239-246.

O'Brien, CA, Pollett, A, Gallinger, S, and Dick, JE (2007). A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. Nature; **445**, 106-110.

Odorico, JS, Kaufman, DS, and Thomson, JA (2001). Multilineage differentiation from human embryonic stem cell lines. Stem Cells; **19**, 193-204.

Olive, KP, and Tuveson, DA (2006). The Use of Targeted Mouse Models For Preclinical Testing of Novel Cancer Therapeutics. Clinical Cancer Research; **12**, 5277-5287.

Palmqvist, L, Glover, CH, Hsu, L, Lu, M, Bossen, B, Piret, JM, Humphries, RK, and Helgason, CD (2005). Correlation of Murine Embryonic Stem Cell Gene Expression Profiles with Functional Measures of Pluripotency. Stem Cells; **23**, 463-480.

Perelygin, AA, Zharkikh, AA, Scherbik, SV, and Brinton, MA (2006). The Mammalian 2'-5' Oligoadenylate Synthetase Gene Family: Evidence for Concerted Evolution of Paralogous Oas1 Genes in Rodentia and Artiodactyla. Journal of Molecular Evolution; **63**, 562-576.

Perry Gardner, H, Ha, SI, Reynolds, C, and Chodosh, LA (2000). The CaM Kinase, Pnck, Is Spatially and Temporally Regulated during Murine Mammary Gland Development and May Identify and Epithelial Cell Subtype Involved in Breast Cancer. Cancer Research; **60**, 5571-5577.

Pesce, M, and Scholer, HR (2001). Oct-4: Gatekeeper in the Beginnings of Mammalian Development. Stem Cells; **19**, 271-278.

Pierce, GB (1974). Neoplasms, differentiations and mutations. American Journal of Pathology; **77**, 103-118.

Pohlers, M, Truss, M, Frede, U, Scholz, A, Strehle, M, Kuban, RJ, Hoffmann, B, Morkel, M, Birchmeier, C, and Hagemeier, C (2005). A Role for E2F6 in the Restriction of Male-Germ-Cell-Specific Gene Expression. Current Biology; **15**, 1051-1057.

Poulos, C, Cheng, L, Zhang, S, Gersell, DJ, and Ulbright, TM (2006). Analysis of Ovarian Teratomas for Isochromosome 12p: Evidence Supporting a Dual

Histogenetic Pathway for Teratomatous Elements. Modern Pathology; **19**, 766-771.

Qin, LX, Beyer, RP, Hudson, FN, Linford, NJ, Morris, DE, and Kerr, KF (2006). Evaluation of Methods for Oligonucleotide Array Data Via Quantitative Real-Time PCR. BMC Bioinformatics; **7**, 23.

Radtke, F, and Clevers, H (2005). Self-Renewal and Cancer of the Gut: Two Sides of a Coin. Science; **307**, 1904-1909.

Rajeevan, MS, Ranamukhaarachchi, DG, Vernon, SD, and Unger, ER (2001a). Use of Real-Time Quantitative PCR to Validate the Results of cDNA Array and Differential Display PCR Technologies. Methods; **25**, 443-451.

Rajeevan, MS, Vernon, SD, Taysavang, N, and Unger, ER (2001b). Validation of Array-Based Gene Expression Profiles by Real-Time (kinetic) RT-PCR. Journal of Molecular Diagnostics; **3**, 26-31.

Ramalho-Santos, M, Yoon, S, Matsuzaki, Y, Mulligan, RC, and Melton, DA (2002). "Stemness": Transcriptional Profiling of Embryonic and Adult Stem Cells. Science; **298**, 597-600.

Rana, TM (2007). Illuminating the Silence: Understanding the Structure and Function of Small RNAs. Nature Reviews Molecular Cell Biology; **8**, 23-36.

Rao, M (2004). Conserved and Divergent Paths that Regulate Self-Renewal in Mouse and Human Embryonic Stem Cells. Dev Biol; **275**, 269-286.

Reed, CC, Waterhouse, A, Kirby, S, Kay, P, Owens, RT, McQuillan, DJ, and Iozzo, RV (2005). Decorin prevents metastatic spreading of breast cancer. Oncogene; **24**, 1104-10.

Reiner, A, Yekutieli, D, and Benjamini, Y (2003). Identifying Differentially Expressed Genes using False Discovery Rate Controlling Procedures. Bioinformatics; **19**, 368-375.

Reis-Filho, JS, Westbury, C, and Pierga, J-Y (2006). The Impact of Expression Profiling on Prognostic and Predictive Testing in Breast Cancer. Journal of Clinical Pathology; **59**, 225-231

Reubinoff, BE, Pera, MF, Fong, C-Y, Trounson, A, and Bongso, A (2000). Embryonic Stem Cell Lines From Human Blastocysts: Somatic Differentiation In Vitro. Nature Biotechnology; **18**, 399-404.

Reya, T, Morrison, SJ, Clarke, MF, and Weissman, IL (2001). Stem cells, cancer, and cancer stem cells. Nature; **414**, 105-111.

Rosenthal, N, and Brown, S (2007). The Mouse Ascending: Perspectives for Human-Disease Models. Nature Cell Biology; **9**, 993-999.

Santamaria, C, Chillon, MC, Garcia-Sanz, Balanzategui, A, Sarasquete, ME, Alcoceba, M, Ramos, F, Bernal, T, Queizan, JA, Penarubia, MJ, Giraldo, P, San Miguel, JF, and Gonzalez, M (2008). The Relevance of Preferentially Expressed Antigen of Melanoma (PRAME) as a Marker of Disease Activity and Prognosis in Acute Promyelocytic Leukaemia. Haematologica; **93**, 1797-1805.

Sato, N, Meijer, L, Skaltsounis, L, Greengard, P, and Brivanlou, AH (2004). Maintenance of Pluripotency in Human and Mouse Embryonic Stem Cells Through Activiation of Wnt Signalling by a Pharmacological GSK-3-Specific Inhibitor. Nature Medicine; **10**, 55-63.

Sato, N, Sanjuan, IM, Heke, M, Uchida, M, Naef, F, and Brivanlou, AH (2003). Molecular Signature of Human Embryonic Stem Cells and Its Comparison With the Mouse. Developmental Biology; **260**, 404-413.

Schiemann, WP, Blobe, GC, Kalumel, DE, Pandeyl, A, and Lodish, HF (2002). Context-specific Effects of Fibulin-5 (DANCE/EVEC on Cell Proliferation, Motility, and Invasion. Journal of Biological Chemistry; **277**, 27367-27377.

Schmittgen, TD, and Livak, KJ (2008). Analyzing real-time PCR data by the comparative Ct method. Nautre Protocols; **3**, 1101-1108.

Schroeder, A, Mueller, O, Stocker, S, Salowsky, R, Leiber, M, Gassmann, M, Lightfoot, S, Menzel, W, Granzow, M, and Ragg, T (2006). The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC Molecular Biology; **7**, 3.

Scohy, S, Gabant, P, Szpirer, C, and Szpirer, J (2000). Identification of an Enhancer and an Alternative Promoter in the First Intron of the Alph-Fetoprotein Gene. Nucleic Acids Research; **28**, 3743-3751.

Seidler, DG, Goldoni, S, Agnew, C, Cardi, C, Thakur, ML, Owens, RT, McQuillan, DJ, and Iozzo, RV (2006). Decorin protein core inhibits in vivo cancer growth and metabolism by hindering epidermal growth factor receptor function and triggering apoptosis via caspase-3 activation. J Biol Chem; **281**, 26408-18.

Sell, S (2004). Stem cell origin of cancer and differentiation therapy. Critical Reviews in Oncology/Hematology; **51**, 1-28.

Shaw, AT, Kirsch, DG, and Jacks, T (2005). Future of Early Detection of Lung Cancer: The Role of Mouse Models. Clinical Cancer Research; **11**, 4999s-5003s.

Shen, H-M, and Pervaiz, S (2006). TNF Receptor Superfamily-Induced Cell Death: Redox-Dependent Execution. The FASEB Journal; **20**, 1589-1598.

Shen, MM, and Leder, P (1992). Leukemia inhibitory factor is expressed by the preimplantation uterus and selectively blocks primitive ectoderm formation in vitro uterus and selectively blocks primitive ectoderm formation in vitro. Proc Natl Acad Sci USA; **89**, 8240-8244.

Shi, W, Wang, H, Pan, G, Geng, Y, Guo, Y, and Pei, D (2006). Regulation of the Pluripotency Marker Rex-1 by Nanog and Sox2. Journal of Biological Chemistry; **281**, 23319-23325.

Singh, SK, Clarke, ID, Terasaki, M, Bonn, VE, Hawkins, C, Squire, JA, and Dirks, PB (2003). Identification of a Cancer Stem Cell in Human Brain Tumours. Cancer Research; **63**, 5821-5828.

Skotheim, RI, Lind, GE, Monni, O, Nesland, JM, Abeler, VM, Fossa, SD, Duale, N, Brunborg, G, Kallioniemi, O, Andrews, PW, and Lothe, RA (2005). Differentiation of Human Embryonal Carcinomas In vitro and In vivo Reveals Expression Profiles Relevant to Normal Development. Cancer Research; **65**, 5588-5598.

Smith, AG, Heath, JK, Donaldson, DD, Wong, GG, Moreau, J, Stahl, M, and Rogers, D (1988). Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. Nature; **336**, 688-670.

Soprano, DR, Soprano, KJ, Wyatt, ML, and Goodman, DS (1988). Induction of the Expression of Retinol-binding Protein and Transthyretin in F9 Embryonal

Carcinoma Cells Differentiated to Embryoid Bodies. Journal of Biological Chemistry; **263**, 17897-17900.

Sorlie, T, Perou, CM, Tibshirani, R, Aas, T, Geisler, S, Johnsen, H, Hastie, T, Eisen, MB, van de Rijn, M, Jeffrey, SS, Thorsen, T, Quist, H, Matese, JC, Brown, PO, Botstein, D, Lonning, PE, and Borresen-Dale, A-L (2001). Gene Expresion Patterns of Breast Carcinomas Distingusih Tumor Subclasses with Clinical Implications. Proc Natl Acad Sci USA; **98**, 10869-10874.

Sperger, JM, Chen, X, Draper, JS, Antosiewicz, JE, Chon, CH, Jones, SB, Brooks, JD, Andrews, PW, Brown, PO, and Thomson, JA (2003). Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors. Proc Nat Acad Sci USA; **100**, 13350-13355.

Stefano, GB, Burrill, JD, Labur, S, Blake, J, and Cadet, P (2005). Regulation of various genes in human leukocytes acutely exposed to morphine: expression microarray analysis. Med Sci Monit; **11**, 35-42.

Stewart, R, Stojkovic, M, and Lako, M (2006). Mechanisms of self-renewal in human embryonic stem cells. European Journal of Cancer; **42**, 1257-1272.

Suarez-Farinas, M, Noggle, S, Heke, M, Hemmati-Brivanlou, A, and Magnasco, MO (2005). Comparing Independent Microarray Studies: The Case of Human Embryonic Stem Cells. BMC Genomics; **6**, 99.

Sugasawa, H, Ichikura, T, Tsujimoto, H, Kinoshita, M, Morita, D, Ono, S, Chochi, K, Tsuda, H, Seki, S, and Mochizuki, H (2008). Prognostic Significance of Expression of CCL5/RANTES Receptors in Patients With Gastric Cancer. Journal of Surgical Oncology; **97**, 445-450.

Sullivan, KM, Bissonette, R, Yanagisawa, H, Hussain, SN, and Davis, EC (2007). Fibulin-5 functions as an endogenous angiogenesis inhibitor. Laboratory Investigation; **87**, 818-827.

Tai, H-H, Tong, M, and Ding, Y (2007). 15-Hydroxyprostaglandin Dehydrogenase (15-PGDH) and Lung Cancer. Prostaglandins Other Lipid Mediat; **83**, 203-208.

Taipale, J, and Beachy, PA (2001). The Hedgehog and Wnt signalling pathways in cancer. Nature; **411**, 349-354.

Tanaka, TS, Kunath, T, Kimber, WL, Jaradat, SA, Stagg, CA, Usuda, M, Yokota, T, Niwa, H, Rossant, J, and Ko, MSH (2002). Gene Expression Profiling of Embryo-Derived STem Cells Reveals Candidate Genes Associated With Pluripotency and Lineage Specificity. Genome Research; **12**, 1921-1928.

Tanno, B, Vitali, R, De Arcangelis, D, Mancini, C, Eleuteri, P, Dominici, C, and Raschella, G (2006). Bim-dependent Apoptosis Follows IGFBP-5 Down-regulation in Neuroblastoma Cells. Biochemical and Biophysical Research Communications; **35**, 547-552.

Theodosiou, A, and Ashworth, A (2002). MAP Kinase Phosphatases. Genome Biology; **3**, 3009.1-3009.10.

Thomas, PD, Campbell, MJ, Kejariwal, A, Mi, H, Karlak, B, Daverman, R, Diemer, K, Muruganujan, A, and Narechania, A (2003a). PANTHER: A Library of Protein Families and Subfamilies Indexed by Function. Genome Research; **13**, 2129-2141.

Thomas, PD, Kejariwal, A, Campbell, MJ, Mi, H, Kiemer, K, Guo, N, Ladunga, I, Ulitsky-Lazareva, B, Muruganujan, A, Rabkin, S, Vandergriff, JA, and Doremieux, O (2003b). PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. Nucleic Acids Research; **31**, 334-341.

Torres, J, and Watt, FM (2008). Nanog Maintains Pluripotency of Mouse Embryonic Stem Cells by Inhibiting NFkB and Cooperating with Stat3. Nature Cell Biology; **10**, 194-201.

Trosko, JE, Chang, C-C, Upham, BL, and Tai, M-H (2005). The Role of Human Adult Stem Cells and Cell-cell Communication in Cancer Chemoprevention and Chemotherapy Strategies. Mutation Research; **591**, 187-197.

Uberall, I, Kolar, Z, Trojenec, R, Berkovcova, J, and Hajduch M (2008). The Status and Role of ErbB Receptors in Human Cancer. Exp Mol Pathol; **84**(2), 79-89.

Vermeulen, L, Sprick, MR, Kemper, K, Stassi, G, and Medema, JP (2008). Cancer Stem Cells - Old Concepts, New Insights. Cell Death and Differentiation; **15**, 947-958.

Wang, H, Rosen, DG, Wang, H, Fuller, GN, Zhang, W, and Liu, J (2006a). Insulin-like Growth Factor-binding Protein 2 and 5 are Differentially Regulated in Ovarian Cancer of Different Histologic Types. Modern Pathology; **19**, 1149-1156.

Wang, Y, Barbacioru, C, Hyland, F, Xiao, W, Hunkapiller, KL, Blake, J, Chan, F, Gonzalez, C, Zhang, L, and Samaha, RR (2006b). Large Scale Real-Time PCR Validation on Gene Expression Measurements from Two Commercial Long-Oligonucleotide Microarrays. BMC Genomics; **7**, 59.

Wei, CL, Miura, T, Robson, P, Lim, S-K, Xu, X-Q, Lee, MY-C, Gupta, S, Stanton, LW, Luo, Y, Schmitt, J, Thies, S, Wang, W, Khrebtukova, I, Zhou, D, Liu, ET, Ruan, YJ, Rao, M, and Lim, B (2005). Transcriptome Profiling of Human and Murine ESCs Identifies Divergent Paths Required to Maintain the Stem Cell Fate. Stem Cells; **23**, 166-185.

Wiese, AH, Auer, J, Lassmann, S, Nahrig, J, Rosenberg, R, Hofler, H, Ruger, R, and Werner, M (2007). Identification of Gene Signatures for Invasive Colorectal Tumour Cells. Cancer Detection and Prevention; **31**, 282-295.

Williams, RL, Hilton, DJ, Pease, S, Willson, TA, Stewart, CL, Gearing, DP, Wagner, EF, Metcalf, D, Nicola, NA, and Gough, NM (1988). Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. Nature; **336**, 684-687.

Wong, M, Ziring, D, Korin, Y, Desai, S, Kim, S, Lin, J, Gjertson, D, Braun, J, Reed, E, and Singh, RR (2008). TNFalpha Blockade in Human Diseases: Mechanisms and Future Directions. Clinical Immunology; **126**, 121-136.

Xi, J, and Yang, Z (2008). Expression of RALDHs (ALDH1As) and CYP26s in Human Tissues and During the Neural Differentiation of P19 Embyronal Carcinoma Stem Cell. Gene Expression Patterns; **8**, 438-442.

Yauk, CL, Berndt, ML, Williams, A, and Douglas, GR (2004). Comprehensive Comparison of Six Microarray Technologies. Nucleic Acids Research; **32**, e124.

Ye, H, Yu, T, Temam, S, Ziober, BL, Wang, J, Schwartz, JL, Mao, L, Wong, DT, and Zhou, X (2008). Transcriptomic Dissection of Tongue Squamous Cell Carcinoma. BMC Genomics; **9**, 69.

Ying, Q-L, Nichols, J, Chambers, I, and Smith, A (2003). BMP Induction of Id Proteins Suppresses Differentiation and Sustains Embryonic Stem Cell Self-Renewal in Collaboration with STAT3. Cell; **115**, 281-292.

Yu, W, Imoto, I, Inoue, J, Onda, M, Emi, M, and Inazawa, J (2007). A Novel Amplification Target, DUSP26, Promotes Anaplastic Thyroid Cancer Cell Growth by Inhibiting p38 MAPK activity. Oncogene; **26**, 1178-1187.

Zhang, M, and Rosen, JM (2006). Stem cells in the Aetiology and Treatment of Cancer. Current Opinion in Genetics and Development; **16**, 60-64.