# LEABHARLANN CHOLÁISTE NA TRÍONÓIDE, BAILE ÁTHA CLIATH Ollscoil Átha Cliath

# TRINITY COLLEGE LIBRARY DUBLIN The University of Dublin

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

#### Copyright statement

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

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

#### Liability statement

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

#### **Access Agreement**

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

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

# Investigating the transcriptional regulation of small RNA expression in Salmonella enterica serovar Typhimurium

A dissertation presented for the degree of Doctor of Philosophy, in the Faculty of Science, University of Dublin, Trinity College

February 2015

### Aoife Colgan

Department of Microbiology,

Moyne Institute of Preventive Medicine,

Trinity College Dublin

TRINITY COLLEGE

1 4 MAY 2015

LIBRARY DUBLIN

#### **Declaration**

I, Aoife Colgan, certify that the experimentation recorded herein represents my own work, unless otherwise stated in the text, and has not been previously presented for a higher degree at this or any other University. This thesis may be lent at the discretion of the Librarian.

Aoife Colgan.

#### **Summary**

Salmonella enterica serovar Typhimurium (S. Typhimurium) is an important foodborne pathogen that causes self-limiting gastroenteritis, or more serious systemic infections in susceptible hosts. S. Typhimurium can infect a wide host range and encounters a series of stressful conditions within various host environments. S. Typhimurium expresses a Type Three Secretion System (TTSS) encoded on a pathogenicity island (SPI1), to mediate invasion of the host intestinal epithelium. Once internalised, S. Typhimurium survives and replicates within the Salmonella containing vacuole (SCV). S. Typhimurium expresses a second TTSS, encoded on a second pathogenicity island (SPI2), to survive the harsh intracellular environment and establish systemic infection.

S. Typhimurium operates a carefully co-ordinated programme of gene expression in response to environmental, spatial and temporal cues in order to survive and establish infection, without incurring fitness costs as a result of inappropriate gene expression. Gene expression is controlled at the transcriptional level by protein factors. An important control point for gene expression also occurs post-transcriptionally. Post-transcriptional gene regulation is often mediated by small non-coding RNAs (sRNAs) which may act in cis or in trans to their target genes. Cis-acting sRNAs often display long regions of base-pair complementarity to their target genes. Trans-acting sRNAs usually affect expression of target genes through short imperfect base-pairing interactions, and often require the RNA chaperone, Hfq. Trans-acting sRNAs control gene expression by a diverse range of mechanisms, but a common feature of sRNA-mediated gene regulation involves a base-pairing interaction in the translation initiation region of the target mRNA, resulting in the inhibition or stimulation of translation. sRNAs may also affect target mRNA stability through promotion or through protection of the target transcript from degradation by cellular ribonucleases. Some sRNAs also directly interact with target proteins.

The transcriptional and post-transcriptional regulatory elements within the cell are intimately connected via complex regulatory networks. This thesis explores the extent of the interconnections between the transcriptional and post-transcriptional regulators, by investigating the protein factors that regulate expression of S. Typhimurium sRNAs. We have used an RNA-seq-based transcriptomic approach to explore the regulons of 18 virulence-associated regulators namely,  $\sigma$ -factors, transcription factors, two-component

systems, an RNA chaperone and an exoribonuclease, under infection-relevant conditions. Our particular focus was the regulation of expression of sRNA genes. We confirmed that a global RNA-seq-based approach can be effectively used for the investigation of bacterial gene regulation, including sRNA-based regulons. We identified 170 sRNA genes which are differentially expressed in one or more of the panel of regulatory mutants, allowing us to generate a transcriptional regulatory network which highlights the complex regulatory interconnections between transcriptional and post-transcriptional regulators. Many of the putative regulatory interactions were further characterised, using a mixture of molecular biology approaches, such as northern blotting, and bioinformatic approaches, such as correlative analysis, binding motif analysis and conservation analysis. We identified 14 sRNAs which are predicted to play important roles in *S*. Typhimurium virulence, based on their patterns of regulation.

Two putative virulence-associated sRNAs, STnc520 and STnc1480, were chosen for detailed mechanistic analysis to elucidate their transcriptional regulation using techniques, such as mutant complementation, Chromatin Immunoprecipitation and quantitative real-time PCR. STnc520 was found to be one of a small number of genes which is directly regulated by the SPI1-encoded transcription factor, SprB. We showed that STnc1480 is targeted by the xenogenic silencing protein H-NS. The transcription factors PhoP and SlyA counteract the repressive effects of H-NS at the STnc1480 promoter. Preliminary steps to elucidate the biological roles of STnc520 and STnc1480 involved transcriptomic approaches, a two-plasmid target validation assay and western immunoblotting. Further experimentation is necessary to validate the mRNA targets of STnc520. STnc1480 post-transcriptionally activates expression of *hilA*, the central regulator of SPI1 gene expression, consistent with an important role for STnc1480 in *S*. Typhimurium pathogenesis.

Taken together, our RNA-seq-based transcriptomic analysis of the regulons of 18 virulence-associated proteins in *S*. Typhimurium will provide a valuable community resource for analysis of gene regulation. Furthermore, our investigation of the regulatory inputs to *S*. Typhimurium sRNAs describes the complex interconnections between transcriptional and post-transcriptional regulatory elements, and is a step towards the elucidation of all molecular regulatory interactions involved in *S*. Typhimurium pathogenesis.

#### Acknowledgements

Firstly, I would like to thank my supervisor, Prof. Jay Hinton, for introducing me to the world of *Salmonella* sRNAs and for giving me the chance to work on such an interesting project. Jay's support, patience and advice have been invaluable over the last four years. A huge thanks to Shabi Srikumar, Kristian Händler, Sathesh Sivasankaran, Magali Hèbrard, Disa Hammarlöf, Rocio Canals and Sian Owen for the help, advice and lots of laughs. Special thanks go to Jane Twohig for always looking after us, making sure we had whatever we needed for experiments, especially in the last year, and for always keeping us stocked up on biscuits and cakes! Thanks doesn't even begin to cover what I owe to Carsten Kröger, who has always been so generous with his time, has taught me so much and who always knows the right Youtube video to cheer me up on a bad science day. I owe you an infinite amount of (perfectly brewed with a timer!!!!) Barry's tea. Best of luck to everyone with what comes next, and best of luck to Jay and the Liverpool Hinton lab.

Thank you to my postgraduate thesis committee Prof. Charles Dorman, Dr. Stephen Smith and Dr. Adrian Bracken, for their advice and help during the last four years. Thanks also to the staff of the Moyne, everyone in the prep room and a special thanks to Stephen and Ronan for brightening up every day. Thanks to Jayne, Caroline and Noreen for keeping us all organised. I'd like to thank Karsten Hokamp for computational support, José Puente, for strains and helpful chats, and Joe Wade, for strains and reminding me about SprB. I wish to thank my undergraduate students Méabh and Michael for their help. Thank you to SFI for funding the project, and to SGM and ASM for providing travel grants.

Thank you to all the past and present post-graduates and postdocs who have made the Moyne such an enjoyable place to spend the last four years. I'd also like to thank members of the Dorman lab, especially Shane, Andrew and Fitzy for sharing strains, reagents and expertise, and Heather for sharing all those things, as well as a shoulder to cry on and constant support. Thanks to half of the best MicroSoc committee, Will and Michael. There will be a piñata when we're all done! Thanks to all my friends, especially Coll, Kat, Meaney, the goofs and goof boys, who always make me laugh no matter what.

A very special thanks to Conor for (mostly) putting up with my appalling time management, for always making sure I'm well-fed and for his never-ending encouragement. Thanks to Anne and Amy for all their support in everything too. Finally thank you to my family, Mam, Dad, Áine and Joe, for always believing in me and giving continuous love, support and help...thank you for everything. And to the littlest members of the family, Oisín and Fiachra, thank you for being the two best distractions an auntie could hope for. A special thanks to Ois for all his help with my "boring work"! None of this would have been possible without you all.

#### **Publications and Conference Communications**

#### **Publications:**

#### Srikumar, S., C. Kröger, M. Hebrard, A. Colgan, M. Beckett, et al (2014)

The intra-macrophage transcriptome of *Salmonella* Typhimurium identifies a genus-specific small RNA required for virulence. **PLOS Pathogens** (In revision).

#### Kröger, C., A. Colgan, S. Srikumar, K. Händler, S.K. Sivasankaran, et al (2013)

An infection-relevant transcriptomic compendium for *Salmonella enterica* serovar Typhiurium. **Cell Host & Microbe 14: 683-695**.

#### Finn, S., K. Händler, O. Condell, A. Colgan, S. Cooney, et al (2013)

ProP is required for the survival of desiccated *Salmonella enterica* serovar Typhimurium cells on a stainless steel surface. **Applied & Environmental Microbiology 79: 4376-4384.** 

# Hebrard, M., C. Kröger, S. Srikumar, A. Colgan, K. Händler & J. C. D. Hinton (2012) sRNAs and the virulence of Salmonella enterica serovar Typhimurium. RNA Biology 9: 437-

sRNAs and the virulence of Salmonella enterica serovar Typhimurium. RNA Biology 9: 437, 445.

#### Kröger, C., S. C. Dillon, A. D. Cameron, K. Papenfort, S.K. Sivasankaran et al (2012)

The transcriptional landscape and small RNAs of *Salmonella enterica* serovar Typhimurium. Proceedings of the National Academy of Sciences of the United States of America 109: E1277-1286

#### **Conference Communications:**

#### 2014 Oral presentation

Dublin Academy of Pathogenomics and Infection Biology (DAPI), University College Dublin, Ireland

#### 2013 Poster presentation

DAPI, Trinity College Dublin, Ireland

#### 2013 Poster presentation

4<sup>th</sup> ASM Conference on Salmonella: The Bacterium, the Host and the Environment, Boston, U. S. A.

#### 2013 Poster presentation

3<sup>rd</sup> Conference on Regulating with RNA in Bacteria, Würzburg, Germany

#### 2012 Poster Presentation

Young Microbiologists Symposium on Microbe Signalling, Organisation and Pathogenesis, University College Cork, Ireland

#### 2012 Poster presentation

DAPI, University College Dublin, Ireland

# Contents

Decla	ration	i
Sumn	nary	ii
Ackn	owledgements	iv
Public	cations and Conference Communications	V
Conte	ents	vi
List o	f Figures	XV
List o	f Tables	cviii
List o	f Abbreviations	.xix
List o	f Appendices	xxii
Chap	ter 1	
Intro	oduction	1
1.1	Salmonella	2
1.1.1	Overview of the Salmonella genus	2
1.1.2	The pathogenesis of <i>S</i> . Typhimurium infection	3
1.1.3	Salmonella pathogenicity island 1 (SPI1)	7
1.1.4	Salmonella pathogenicity island 2 (SPI2)	8
1.2	Transcriptional Regulation	10
1.2.1	Methods of transcriptional regulation	10
1.2.2	Dam	17
1.2.3	HilD	18
1.2.4	FliZ	20

1.2.5	BarA/SirA
1.2.6	Fur
1.2.7	$\sigma^{54}$
1.2.8	$\sigma^{38}$
1.2.9	$\sigma^{24}$
1.2.10	PhoB/R
1.2.11	SlyA
1.2.12	SsrA/B
1.2.13	PhoP/Q
1.2.14	OmpR/EnvZ
1.2.15	FNR
1.3	Small non-coding RNAs (sRNAs)
1.3.1	Overview
1.3.2	Cellular functions of <i>trans</i> -acting sRNAs
1.3.3	Hfq47
1.3.4	PNPase49
1.4	Aims of the study50
Chapt	er 2
Mate	rials and Methods52
2.1	Chemicals and reagents53
2.2	General microbiological techniques53
2 2 1	Madia 52

2.2.2	Culture conditions	53
2.2.3	Ectopic expression analysis	54
2.2.4	Motility assay	54
2.2.5	Antibiotics	55
2.2.6	Maintenance of bacterial stocks	55
2.3	General molecular techniques	61
2.3.1	Preparation of chromosomal S. Typhimurium 4/74 DNA for use in PCR	61
2.3.2	Polymerase chain reaction (PCR)	61
2.3.3	Agarose gel electrophoresis	62
2.3.4	Restriction digestion	62
2.3.5	Purification of PCR products	63
2.3.6	Extraction of DNA from agarose gels	63
2.3.7	Plasmid preparations	63
2.4	Preparation and transformation of competent bacteria	65
2.4.1	Preparation of electro-competent cells and electroporation	65
2.4.2	Preparation and transformation of chemically competent bacteria	65
2.5	Genetic manipulations	66
2.5.1	Generation of deletion mutants in S. Typhimurium 4/74	66
2.5.2	Epitope tagging	67
2.5.3	P22 phage transduction	68
2.5.4	Removal of antibiotic resistance cassettes	68
2.5.5	Sequence and Ligation Independent Cloning (SLIC)	69
2.6	Analysis of RNA	70

2.6.1	Extraction of RNA from S. Typhimurium 4/74	.70
2.6.2	DNase I digestion	.71
2.6.3	Reverse transcription	.72
2.6.4	Quantitative real-time PCR	.72
2.6.5	Generating Digoxigenin-labelled riboprobes	.75
2.6.6	Northern blotting	.75
2.6.7	cDNA library preparation for RNA-seq	.77
2.6.8	RNA-seq	.77
2.6.9	Mapping of RNA-seq data and differential expression analysis	.81
2.6.10	Analysis of RNA stability	.81
2.7	Analysis of proteins	.82
2.7.1	Preparation of whole cell lysates for analysis of cellular proteins	.82
2.7.2	Preparation of culture supernatants for analysis of secreted proteins	.82
2.7.3	SDS polyacrylamide electrophoresis (SDS-PAGE)	.83
2.7.4	Coomassie staining of polyacrylamide gels	.83
2.7.5	Western immunoblotting	.84
2.8	Chromatin immunoprecipitation (ChIP)	.85
2.8.1	Preparation of cross-linked lysates	.85
2.8.2	Immunoprecipitation	.86
2.8.3	Washing the Protein G agarose beads and elution of DNA	.86
2.8.4	Reversal of cross-links	.87
2.8.5	DNA extraction and analysis	.87
2.9	GFP-based two-plasmid validation assay for sRNA targets	.88

2.9.1	Plasmids	. 88
2.9.2	The two-plasmid target validation assay	. 88
2.10	Bioinformatic analysis	.89
2.10.1	Calculation of Pearson Correlation Coefficient	. 89
2.10.2	Generation of position-specific scoring matrix (PSSM)	.89
2.10.3	List of online tools and programmes used for bioinformatic analysis in this stu	ıdy
		.89
Chapt	er 3	
	al analysis of regulatory inputs on Salmonella Typhimurium sma	
RNA	S	91
3.1	Introduction	.92
3.1.1	Overview of the study	.92
3.1.2	RNA-seq	.96
3.2	Results	98
3.2.1	Reproducibility of RNA-seq data	98
3.2.2	Construction and sequencing of chromosomal deletion mutants	00
3.2.3	Identification of previously determined regulons validates the RNA-seq approa	ach
	1	05
3.2.3.1	The Fur regulon1	09
3.2.3.2	The $\sigma^{54}$ regulon	10
3.2.3.3	The Hfq regulon1	11
2224	The PNPase regulon1	1.1

3.2.3.5 The SlyA reg	ulon	112
3.2.3.6 The SsrA reg	ulon	113
3.2.3.7 The SsrB regi	ılon	114
3.2.3.8 The OmpR re	gulon	114
3.2.3.9 The FNR regu	ılon	115
	the RNA-seq approach for detectin	
3.2.5 Confirmation	of the regulation of expression of n	new S. Typhimurium sRNAs . 121
3.2.6 Regulatory in	puts on S. Typhimurium expression	125
3.2.7 Global classif	ication of differentially expressed s	RNAs131
3.2.8 Motif analysis	s of differentially expressed sRNAs	134
	ne Fur consensus binding motif and	
	e SsrB consensus binding motif an	
	ne $\sigma^{38}$ consensus binding motif and	
3.2.9 Transcription	al network analysis	142
3.2.10 Identification	of SPI1 and SPI2-like sRNAs	146
3.2.10.1 SPI1-like sR	NAs	147
3.2.10.2 SPI2-like sR	NAs	147
3.2.11 Conservation	of SPI-like sRNAs	151
3.2.11.1 Conservation	of the SPI1-like sRNAs	151
3 2 11 2 Conservation	of the SPI2-like sRNAs	152

3.2.12	2 Investigation of the virulence-associated phenotype of SPI1- and SPI2-li	
3.2.13	3 Analysis of the sRNA regulon of Hfq	155
3.2.14	4 Analysis of the sRNA regulon of PNPase	159
3.3	Discussion	162
3.3.1	Regulation of transcription of S. Typhimurium sRNAs	162
3.3.2	Regulatory families of sRNA genes	163
3.3.3	Regulation of the steady-state levels of S. Typhimurium sRNAs	164
3.3.4	The benefits and limitations of and RNA-seq-based approach for the involved of bacterial regulons	
	stigating the transcriptional regulation of the SPI1-like	
Inves	stigating the transcriptional regulation of the SPI1-like	170
Inves	stigating the transcriptional regulation of the SPI1-like	170
Inves	stigating the transcriptional regulation of the SPI1-like	170
Inves STnc 4.1 4.1.1	stigating the transcriptional regulation of the SPI1-like	170171
Inves STnc 4.1 4.1.1	stigating the transcriptional regulation of the SPI1-like c520	170171171173
Inves STnc 4.1 4.1.1 4.1.2	SPI1-associated sRNAs  SprB	170171171173
Inves STnc 4.1 4.1.1 4.1.2 4.1.3	SPI1-associated sRNAs  SprB  SPI11	170171173174175
Inves STnc 4.1 4.1.1 4.1.2 4.1.3 4.1.4	SPI1-associated sRNAs  SprB  SPI11  Formate metabolism	170171173174175176
Inves STnc 4.1 4.1.1 4.1.2 4.1.3 4.1.4 4.1.5	SPI1-associated sRNAs  SprB  SPI11  Formate metabolism  Sulphur metabolism	170171173174175176178
Inves STnc 4.1 4.1.1 4.1.2 4.1.3 4.1.4 4.1.5 4.1.6	SPI1-associated sRNAs  SprB  SPI11  Formate metabolism  Sulphur metabolism  Copper homeostasis	170171173174175176178

Inves	tigating the transcriptional regulation of STnc1480220
Chapto	er 5
4.3.2	The role of STnc520 within the SprB regulon
4.3.1	The transcriptional regulation of STnc520
4.3	Discussion
4.2.13	209
	Biocomputational prediction of STnc520 binding sites for candidate target genes
4.2.12	Biocomputational prediction of STnc520 target binding sites
4.2.11	Investigation of the mRNA targets of STnc520 using RNA-seq-based technology 201
	Investigation of the interaction between STnc520 and putative target mRNAs . 199
1016	
4.2.9	Investigation of mRNA targets of STnc520 using microarray-based technology
4.2.8	The role of STnc520 in virulence
4.2.7	Conservation of the STnc520 promoter region
4.2.6.2	SprB binds the STnc520 promoter <i>in vivo</i>
4.2.6.1	Construction of a 4/74 sprB-FLAG strain
4.2.6	Analysis of SprB binding to the STnc520 promoter
4.2.5	SprB is the SPI1-encoded transcription factor necessary for STnc520 expression
	expression analysis
4.2.4	Investigation of the role of HilD in regulating STnc520 expression by exogenous
4.2.3	STnc520 shows a SPI1-like pattern of expression

5.1	Introduction	221
5.1.1	Transcriptional silencing and counter-silencing	221
5.1.2	Trans-acting sRNAs that activate gene expression	225
5.2.	Results	226
5.2.1	Characterisation of STnc1480	226
5.2.2	Predicted structure and conservation of STnc1480	229
5.2.3	STnc1480 has a SPI2-like pattern of expression	232
5.2.4	STnc1480 requires PhoP and SlyA for expression	235
5.2.5	Silencing of STnc1480 by H-NS under non-inducing conditions	237
5.2.6	Characterisation of the H-NS-mediated silencing of STnc1480 transcription2	239
5.2.7	Binding of PhoP to the STnc1480 promoter	242
5.2.8	ChIP analysis of PhoP binding in a <i>hns</i> mutant background	245
5.2.9	Binding of SlyA to the STnc1480 promoter	248
5.2.10	The role of STnc1480	251
5.2.11	STnc1480 regulation of <i>hilA</i> expression	254
5.2.12	The effect of STnc1480 on Type Three Secretion systems	258
5.3	Discussion	263
5.3.1	Transcriptional regulation of STnc1480	263
5.3.2	STnc1480 and S. Typhimurium virulence	264
Chapte	er 6	
Gene	ral Discussion2	68
6.1	Importance of this study2	269

6.1.1	Context of the study	269
6.1.2	Global analysis of the transcriptional regulation of sRNAs	270
6.1.3	STnc520	273
6.1.4	STnc1480	273
6.1.5	Global analysis of the post-transcriptional regulation of sRNAs	275
6.2	Future prospects	275
6.3	Concluding remarks	277
Biblic	ography	279
List o	of Figures	

Figure	Page Page
1.1	Pathogenesis of S. Typhimurium6
1.2	Key mechanisms involved in regulating the initiation of bacterial transcription 16
1.3	Main virulence-associated regulatory networks in S. Typhimurium
1.4	Main mechanisms of post-transcriptional regulation by Hfq-bound <i>trans</i> -acting sRNAs
2.1	Absolute expression in TPM of the <i>hemX</i> gene
3.1	Workflow of general experimental procedures for investigating the regulatory inputs on sRNA genes
3.2	BLAST atlas of the S. Typhimurium SL1344 genome showing the location and
	conservation of genes important for this study95

High correlations of RNA-seq data from technical and independent biological replicates	3.3
Confirmation of 19 chromosomal deletions by RNA-seq	3.4
Validation of RNA-seq of mutant strains as a method of investigating Salmonella regulons	3.5
Validation of new regulatory inputs on sRNAs by northern blot	3.6
Global analysis of regulatory inputs on S. Typhimurium sRNA genes	3.7
Global classification of sRNAs that were differentially expressed in at least one regulatory mutant	3.8
Motif analyses of differentially expressed sRNA promoters	3.9
Transcriptional regulatory network of <i>S.</i> Typhimurium sRNAs	3.10
Cluster analyses of all protein-coding and sRNA genes based on their patterns of relative expression in 15 regulatory mutants	3.11
Conservation analyses of SPI1-like and SPI2-like sRNAs	3.12
Analysis of the sRNA regulon of Hfq	3.13
Analysis of the PNPase sRNA regulon	3.14
Chromosomal location and stability of STnc520	4.1
Predicted structure and conservation of STnc520	4.2
STnc520 has a SPI1-like pattern of expression and regulation	4.3
Analysis of STnc520 expression following ectopic expression of HilD	4.4
Expression of STnc520 in isogenic mutants of the key SPI1-encoded and SPI1-associated transcription factors	4.5
Construction and functional activity of 4/74 sprB-FLAG	4.6
SprB directly binds to the STnc520 promoter region	4.7

4.8	Multiple alignment of the STnc520 promoter sequence
4.9	Investigation of the role of STnc520 on SPI1 secretion
4.10	Use of the two-plasmid system to investigate putative mRNA targets of STnc520
4.11	Investigation of the mRNA targets of STnc520 using RNA-seq
4.12	Predicted interactions between STnc520 and putative target mRNAs212
4.13	Model of the transcriptional regulation and putative role of STnc520219
5.1	Chromosomal location and stability of STnc1480
5.2	Conservation and predicted structure of STnc1480
5.3	STnc1480 has a SPI2-like pattern of expression and regulation
5.4	Analysis of STnc1480 expression following ectopic expression of PhoP and SlyA
5.5	H-NS binds the STnc1480 promoter and silences STnc1480 transcription238
5.6	Characterisation of H-NS-mediated silencing of STnc1480 transcription241
5.7	Epitope tagging of PhoP and investigation of PhoP binding to the STnc1480 promoter in a wild-type background
5.8	Investigation of PhoP binding to the STnc1480 promoter in an <i>hns</i> mutant background
5.9	FLAG-tagging the <i>slyA</i> gene
5.10	The hilA mRNA 5' UTR is a STnc1480 target
5.11	STnc1480 activation of HilA in S. Typhimurium
5.12	The effect of STnc1480 on secreted levels of FliC
5.13	Model of the transcriptional regulation and role of STnc1480

# **List of Tables**

Table	Title Page
2.1	Liquid media recipes
2.2	Bacterial strains and plasmids
2.3	Strains and growth conditions used for HiSeq RNA-seq79
2.4	Antibodies used for western immunoblot analysis
3.1	Comparison of RNA-seq-based transcriptomic data with published regulatory interactions between TFs and sRNAs
3.2	SPI1-like and SPI2-like sRNAs
4.1	Genes showing significant (> 2-fold) and reproducible (p < 0.05) differential expression following pulse over-expression of STnc520 and microarray analysis
4.2	Interpretation of RNA-seq-based STnc520 target-hunting data
4.3	RNA-seq data showing differential expression of genes in $\Delta STnc520$ or following pulse over-expression of STnc520
4.4	Candidate STnc520 target genes based on bioinformatic analysis

#### List of Abbreviations

3' UTR Three prime untranslated region

4/74 Salmonella enterica serovar Typhimurium strain 4/74

5' UTR Five prime untranslated region

Amp Ampicillin

ATR Acid tolerance response

bp Basepairs

cDNA complementary DNA

CDS coding sequence

CFU colony forming units

ChIP Chromatin immunoprecipitation

CLR Context likelihood of relatedness

Cm Chloramphenicol

Dig Digoxigenin

DMSO Dimethyl sulfoxide

DNA Deoxyribonucleic acid

dNTP deoxyribonucleotide

EBP Enhancer binding protein

EDTA Ethylenediaminetetraacetic acid

EMSA Electrophoretic mobility shift assay

ESP Early stationary phase

ESR Envelope stress response

FLP Flippase recombinase

FRT Flippase recognition target

FRUIT Flexible recombineering using integration of thyA

GFP Green fluorescent protein

HGT Horizontal gene transfer

HRP Horseradish peroxidase

HRT Horizontal regulatory transfer

H-T-H Helix-turn-helix

IGB Integrated Genome Browser

IGR Intergenic region

inSPI2 SPI2-inducing

Kan Kanamycin

kb Kilobases

LEP Late exponential phase

LPS Lipopolysaccharide

LSP Late stationary phase

MCS Multiple cloning site

MEP Mid exponential phase

MMA minimal medium A

mRNA Messenger RNA

NAP Nucleoid associated protein

nt Nucleotide

NTS Non-typhoidal Salmonella

OD<sub>600</sub> Optical density at 600 nm

OMP Outer membrane protein

ORF Open reading frame

PBS Phosphate buffered saline

PCN Phosphate, Carbon, Nitrogen medium

PCR Polymerase chain reaction

PSSM Position-specific scoring matrix

qPCR quantitative real-time PCR

RBS Ribosome binding site

Rif Rifampicin

RNA Ribonucleic acid

RNAP RNA polymerase

RNase Ribonuclease

RNA-seq High-throughput cDNA sequencing

RNS Reactive nitrogen species

ROS Reactive oxygen species

RPKM Reads per kilobase per million

rpm Revolutions per minute

RT Reverse transcripase/transcripion

SBS Sequencing by synthesis

SCFA Short chain fatty acid

SCV Salmonella containing vacuole

SDS- Sodium dodecyl sulphate poly-acrylamide

PAGE electrophoresis

SIF Salmonella induced filaments

SLIC Sequence and ligation independent cloning

SP Spacious phagosome

SPI Salmonella pathogenicity island

sRNA small non-coding RNA

ST19 Sequence Type 19

STncXXX Salmonella Typhimurium non-coding

TAE Tris acetate ethylenediaminetetraacetic acid

TCS Two component system

TF Transcription factor

TIR Translation initiaiton region

TMAO Trimethylamine N-oxide

TPM Transcripts per million

TraDIS Transposon directed insertion site sequencing

tRNA Transfer RNA

TSS Transcription start site

TTSS Type three secretion system

UV Ultraviolet

WT Wild-type

#### **List of Appendices (contents of accompanying CD)**

- I TPM of all genes in 19 regulatory mutants
- II Differentially expressed protein-coding genes (HiSeq)
- III Differentially expressed sRNA genes (HiSeq)
- IV Regulon comparisons for Figure 3.5
- V STnc520 target-hunting results (MiSeq)
- VI Predicted STnc520 target binding sites TargetRNA2, IntaRNA, CopraRNA
- VII pXG10 (sf) fusions
- VIII Growth and activity of fnr mutant in MMA
- IX Oligonucleotide table

# Chapter 1 Introduction

#### 1.1 Salmonella

#### 1.1.1 Overview of the Salmonella genus

Salmonellae are Gram negative, rod-shaped, facultative anaerobic members of the bacterial family Enterobacteriaceae (Fabrega & Vila, 2013). The Salmonella genus diverged from a common ancestor with Escherichia coli (E. coli) between 100 and 150 million years ago (Sabbagh et al., 2010). There are two recognised species within the Salmonella genus, Salmonella bongori and Salmonella enterica. Salmonella enterica is further subdivided into six subspecies, enterica (I), salamae (II), arizonae (IIIa), diarizonae (IIIb), houtenae (IV) and indica (VI). Salmonellae have a broad host range but Salmonella enterica subspecies enterica is most commonly associated with infection of humans and other warm-blooded hosts, while the other S. enterica subspecies and S. bongori are more commonly associated with infection of cold-blooded vertebrates or are isolated from environmental sources (Desai et al., 2013). There are currently 2579 serovars within the entire Salmonella genus, and 1531 of these serovars belong to the subspecies enterica (Grimont & Weill, 2007). For simplicity in this study, serovars of Salmonella enterica subsp. enterica will be referred to using the serovar name, for example, Salmonella Typhimurium (S. Typhimurium).

The two main clinical manifestations associated with *Salmonella enterica* subsp. *enterica* are gastroenteritis and enteric fever, also known as Typhoid fever. Typhoid fever and the clinically indistinguishable illness, Paratyphoid fever, are caused by the human host-restricted serovars, Typhi and Paratyphi (McClelland *et al.*, 2004). The early stages of Typhoid may go unnoticed, as typhoidal *Salmonella* strains do not produce a pro-inflammatory response resulting in gastroenteritis, and thus the infection might not be treated. At later stages, the symptoms of infection include fever and associated malaise, as well as a progression to systemic infection or the possibility of entering a carrier state (Dougan & Baker, 2014). *S.* Typhi and *S.* Paratyphi are typically transmitted through faecal contamination of water sources or food, and Typhoid is currently more common in the developing world. It is estimated that there are greater than 20 million new cases of Typhoid worldwide annually, resulting in approximately 200,000 deaths per year (Crump & Mintz, 2010).

Non-typhoidal Salmonella (NTS) strains, by contrast, do produce an inflammatory response upon invasion of host epithelial cells. The pro-inflammatory response leads to a self-limiting gastroenteritis, the symptoms of which include diarrhoea with or without blood, abdominal cramps, nausea and fever. S. Typhimurium and S. Enteritidis are the most commonly isolated NTS serovars in clinical settings. A typical NTS infection in a developed country lasts less than a week and rarely requires hospitalisation or treatment, however more complicated sequelae may develop in immunocompromised hosts (Fabrega & Vila, 2013). Again, NTS strains are most commonly transmitted through ingestion of contaminated food and water. As S. enterica is also a zoonotic pathogen, infection of food-producing animals can lead to the dissemination of the pathogen throughout the food-processing industry. In the developed world there are an estimated 93.8 million incidences of Salmonellosis caused by NTS strains, predominantly S. Typhimurium Sequence Type 19 (ST19), which result in 155,000 deaths each year (a mortality rate of less than 0.2%) (Majowicz et al., 2010). In developing countries, however, the mortality rate associated with NTS sequence types is typically between 20-25% due to the combination of underlying conditions, such as HIV, malaria and malnutrition, as well as the emergence of more invasive NTS strains (Feasey et al., 2012).

S. Typhimurium ST19 is an ideal model to study the disease progression of both clinical manifestations of *Salmonella* infection as S. Typhimurium produces gastroenteritis in human hosts and a Typhoid-like fever in mice (Sabbagh et al., 2010).

#### 1.1.2 The pathogenesis of S. Typhimurium infection

The infection process of *S*. Typhimurium involves a series of stressful conditions for the bacterium (Figure 1.1). Upon ingestion from a contaminated food or water source, the bacterium first experiences a temperature upshift to physiological conditions. The acidic environment of the stomach also represents a stressful environment and *Salmonella* induces an acid tolerance response (ATR) to resist severe acid shock, and to prepare the bacterial cells for further acidic conditions during intracellular infection (Foster & Hall, 1990, Foster, 1991). The bacteria enter and descend through the small intestine to reach the preferential site for invasion, the ileum. The bacterium encounters increased osmolarity and anaerobiosis, as well as the host intestinal microbiota during this journey (Hebrard *et al.*, 2011). A functional motility and chemotaxis system is a prerequisite for

Salmonella to approach and encounter the intestinal epithelium (Jones et al., 1981). Salmonella is peritrichously flagellated and the flagellar system requires more than 50 genes, from approximately 17 operons (Chilcott & Hughes, 2000). Salmonella transverses the epithelial mucus layer, avoiding destruction by elements of the host innate immune system, such as bile and antimicrobial peptides (Haraga et al., 2008), and directly adheres to and invades non-phagocytic intestinal epithelial cells, however microfold (M) cells of the Peyer's patches are the preferred target host cells (Fabrega & Vila, 2013). Intimate attachment of the bacteria to the host epithelial cell is necessary for induction of invasion-associated genes (Jones et al., 1981). Generation of close contact interactions between the bacterial and host cells is mediated by Salmonella surface appendages, such as Type 1 fimbriae, curli fimbriae and non-fimbrial adhesins (Wagner & Hensel, 2011). Adherence of Salmonella to the apical side of the intestinal epithelial cell induces significant host cell cytoskeletal rearrangement and membrane ruffling that results in engulfment of the bacterial cell (Finlay et al., 1991). This process is mediated by translocation of effector proteins through a Type Three Secretion System (TTSS), encoded on a pathogenicity island known as Salmonella Pathogenicity Island 1 (SPI1) (Haraga et al., 2008). Detection of the invading bacteria by components of the host immune system leads to a pro-inflammatory immune response, resulting in the production of the symptoms of Salmonellosis and gastroenteritis (Thiennimitr et al., 2012). A portion of bacteria remain in the intestinal lumen of the host, rather than invade the epithelial layer, and benefit from the induction of inflammation of the intestine. Antimicrobial peptides and reactive oxygen species (ROS) kill non-pathogenic commensals of the gut, but resistance mechanisms allow Salmonella to out-compete the host microbiota and thrive in the inflamed intestine (Thiennimitr et al., 2012). As an example, gut microbiota produce the compound hydrogen sulfide (H<sub>2</sub>S), which is converted to thiosulphate in the gut. Neutrophils, recruited to the gut lumen to protect against invading bacteria, produce ROS that oxidises thiosulphate and produces tetrathionate. In contrast to the gut microbiota, Salmonella can use tetrathionate as an alternative electron acceptor in anaerobic respiration. This process is more energy-efficient than fermentation, thus providing Salmonella with a growth advantage over the fermentative gut bacteria (Winter et al., 2010).

The fraction of bacteria that become internalised in epithelial cells reside intracellularly in membrane-derived vesicles known as spacious phagosomes (SP). These phagosomes

shrink to form an adherent membrane around one or more bacteria and this compartment is known as the Salmonella containing vacuole (SCV) (Haraga et al., 2008). Using effector proteins, secreted through a second TTSS encoded on Salmonella Pathogenicity Island 2 (SPI2), and accessory proteins, Salmonella can survive and replicate within the SCV (Fabrega & Vila, 2013). Formation of the SCV involves altering the host cell endocytic trafficking pathway to avoid normal phagosome maturation and fusion with lysosomes (Rathman et al., 1997). During SCV maturation, Salmonella encounters magnesium, phosphate and iron starvation and a further reduction in pH (Hebrard et al., 2011). Salmonella resists killing by host innate immune factors and is likely to acquire molecules necessary for bacterial replication by positioning of the SCV close to the Golgi apparatus (Salcedo & Holden, 2003). Salmonella also induces the formation of tubular filamentous structures known as Salmonella Induced Filaments (SIFs), which extend from the SCV and form complex networks throughout the cell. SIFs exist as single or double membrane stable structures that maintain the SCV in a Golgi-associated and juxtanuclear position, while supplying the SCV with endocytosed nutrients (Krieger et al., 2014). In addition to this traditional view of vacuole-contained bacteria, a significant portion of the internalised bacteria within intra-epithelial cell SCVs lyse their phagosome and escape to the host cell cytosol. Cytosolic bacteria evade destruction by autophagic mechanisms of the host epithelial cell, and survive to hyper-replicate later during infection (Knodler et al., 2014).

A portion of the intact SCVs then transcytose to the basolateral membrane of the epithelial cell and, once across the intestinal epithelium, the bacteria are engulfed by phagocytic cells, primarily macrophages, leading to bacterial internalisation within SCVs again. *Salmonellae* within intra-macrophage SCVs trigger similar host cell pathways to those triggered by intra-epithelial cell SCVs, avoiding phagosome maturation and allowing the bacteria to proliferate (Fabrega & Vila, 2013). The intra-epithelial cell SCV and the intra-macrophage SCV share a number of common characteristics as well as some key differences. Nutrient starvation and acidic pH is a feature of both intracellular vacuoles, but the intra-macrophage environment induces expression of systems involved in the resistance to oxidative and nitrosative stress, which reflects the more bactericidal nature of the macrophage environment (Hautefort *et al.*, 2008). The migration of infected macrophages allows for dissemination of *S*. Typhimurium, via the bloodstream, to organs such as the spleen and liver in mice and immunocompromised hosts (Haraga et al., 2008).

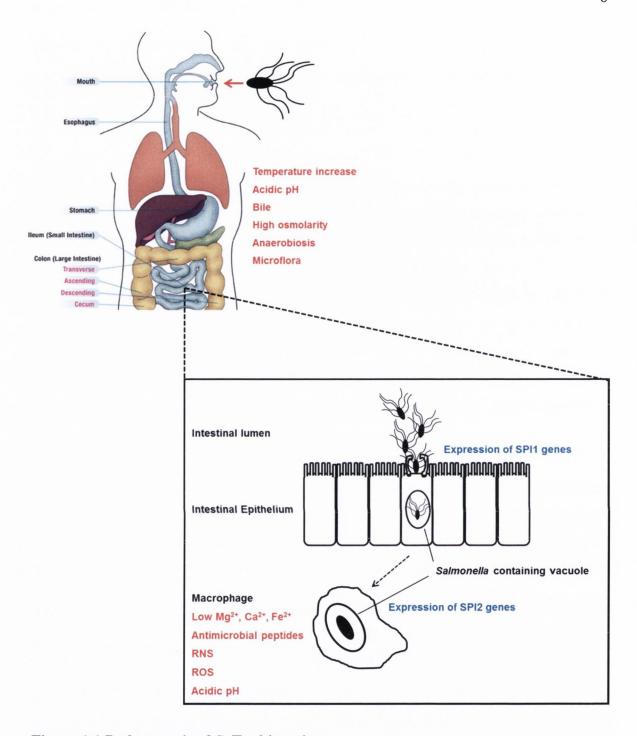


Figure 1.1 Pathogenesis of S. Typhimurium

S. Typhimurium is ingested from contaminated food or water sources. While descending through the gastrointestinal tract S. Typhimurium encounters a variety of stressful conditions (labelled in red). Zoomed in panel: S. Typhimurium invades epithelial cells of the ileum. Environmental signals within the gut induce expression of the SPI1 TTSS, which injects effector proteins into the host epithelial cell that trigger intestinal inflammation and bacterial endocytosis. S. Typhimurium survives and replicates intracellularly within a Salmonella containing vacuole (SCV). In susceptible hosts, S. Typhimurium becomes engulfed by macrophages and survives and replicates within the SCV, leading to bacterial dissemination and systemic infection. Conditions encountered by S. Typhimurium within the macrophage SCV are labelled in red. These environmental conditions lead to the induction of the SPI2 TTSS, which injects effector proteins, necessary for the intracellular lifestyle, into the macrophage cytosol.

#### 1.1.3 Salmonella Pathogenicity Island 1 (SPI1)

The *S.* Typhimurium genome contains 13 pathogenicity islands, or SPIs (Kröger et al., 2012). SPIs 1-5 have been clearly shown to play a direct role in *S.* Typhimurium virulence (Fabrega & Vila, 2013) and SPI1 and SPI2 are the best studied pathogenicity islands as a result of the key role they play in the infectious process of *S.* Typhimurium. SPI1 is a 40 kb cluster of *Salmonella*-specific DNA that encodes all the components necessary to form a functional TTSS apparatus, the effectors which are translocated into the host cytoplasm by the TTSS, as well as regulators of SPI1 gene expression (Ellermeier & Slauch, 2007). TTSSs are widespread among Gram negative pathogenic bacteria as a method of injecting bacterial effector proteins into host cells to mediate bacterial invasion or co-option of host cell function (Galan, 2001). Expression of SPI1-encoded genes is induced in the intestinal epithelium, following adherence of the bacterium to the host cell, and SPI1 genes mediate the host cell actin cytoskeleton rearrangement, as well as engulfment and internalisation of the bacteria (Zhou & Galan, 2001).

The SPI1 TTSS is embedded in the inner and outer bacterial membranes, spans the periplasmic space and contains a needle-like structure that extends into the extracellular environment and makes contact with the host cell (Schraidt & Marlovits, 2011). The components of the SPI1apparatus are organised into distinct gene clusters. In general terms the prg/org and inv/spa clusters encode the needle complex, while the sic/sip operon encodes the translocon that embeds into the host cell membrane, as well as some effector proteins (Ellermeier & Slauch, 2007). The apparatus is composed of a multimeric ring-shaped base structure which spans the inner and outer bacterial membranes. An inner ring comprises the PrgH and PrgK proteins, while the InvG protein forms an outer ring and neck region of the base structure (Schraidt et al., 2010). A needle-like structure extends from the base, comprising the PrgI protein (Kimbrough & Miller, 2000) and inner rod, PrgJ. The inner rod acts as a molecular ruler to modulate the length of the needle complex and to switch TTSS substrate (Lefebre & Galan, 2014). When contact with the host cell is made, the translocon, which is a pore-like complex consisting of the SipB, SipC and SipD proteins, is secreted via the TTSS and embeds in the host cell membrane (Collazo & Galan, 1997).

Three SPI1-translocated effector proteins, SopE, SopE2 and SopB, which are encoded elsewhere on the chromosome, direct host cell cytoskeleton re-modelling by activating the host Rho GTPases Cdc42, Rac1 and RhoG, leading to membrane ruffling and bacterial engulfment by macropinocytosis (Galan & Zhou, 2000, Stender *et al.*, 2000). The SPI1-encoded SipA is also translocated via the SPI1 TTSS and contributes to bacterial engulfment by stabilising actin at the site for bacterial entry, and increasing the membrane extrusions at this site (Zhou *et al.*, 1999). In addition to its role as part of the translocon complex, SipC nucleates actin filaments which aids in membrane ruffling and increases bacterial uptake (Galan & Zhou, 2000). Three additional effector proteins, SopD, SopA and the SPI1-encoded IacP contribute to intestinal inflammation (Fabrega & Vila, 2013). SPI1 encodes a number of chaperone proteins that function to stabilise and target their substrate proteins for translocation (Fabrega & Vila, 2013). Correct folding and functioning of SPI1-secreted effector proteins requires the disulphide oxidoreductase, DsbA, which is involved in the formation of periplasmic disulphide bonds (Ellermeier & Slauch, 2004).

Interestingly, the host cell actin cytoskeleton resumes its normal conformation shortly after infection, despite the presence of intracellular bacteria within SCVs in the epithelial cell. This feature has been attributed to the SPI1-encoded and secreted SptP protein, which acts as a GTPase activating protein, stimulating the intrinsic GTPase activity of the Rho GTPase proteins, resulting in their conversion to their inactive GDP-bound form. The SptP-mediated reversal of actin cytoskeletal changes has been proposed to prevent excessive damage to the host cell that could cause host cell apoptosis and prevent the survival and replication of bacterial cells (Galan & Zhou, 2000).

#### 1.1.4 Salmonella Pathogenicity Island 2 (SPI2)

A second TTSS encoded on a second pathogenicity island, SPI2, was identified by signature-tagged transposon mutagenesis. Mutants of the genes encoded within this cluster are attenuated for murine infection via the oral and intraperitoneal (i.p.) routes, indicating that the SPI2 TTSS plays an important role after the colonisation and invasion of epithelial cells (Shea *et al.*, 1996). The pathogenicity island was initially found to be 40 kb in length but subsequent analysis determined that the island is mosaic in nature and

consists of 2 sections. The first section is a 15 kb region that encodes genes necessary for anaerobic respiration using tetrathionate as a terminal electron acceptor and other genes of unknown function. The second 25 kb region of the island encodes the TTSS and accessory proteins (Hensel, 2000). Genes encoded within the 25 kb region were named according to their predicted or confirmed functions. Genes encoding structural proteins of the TTSS apparatus were named ssa, genes encoding effector proteins were named sse, genes encoding chaperone proteins were named ssc and genes encoding regulatory proteins were named ssr (Hensel, 2000). In contrast to the SPI1-encoded TTSS, the function of the SPI2-encoded TTSS is to mediate Salmonella survival within the intracellular compartment. SPI2-secreted effector proteins are translocated into the host cell cytoplasm across the membrane of the SCV to manipulate host cell function and to allow the bacteria to survive and replicate intracellularly (Abrahams & Hensel, 2006). The SPI2 locus is not found in the phylogenetically older Salmonella bongori species, indicating that Salmonella enterica acquired the ability to survive intracellularly and cause systemic infection since the divergence of the two species (Ochman & Groisman, 1996).

Salmonellae can survive and replicate in intracellular compartments in many host cell types, particularly epithelial cells and macrophages. The macrophage environment appears to be the more hostile intracellular environment due to the presence of antimicrobial peptides, acidic pH and the production of ROS, RNS and their reactive intermediates (Hautefort et al., 2008). The specific functions of many SPI2-secreted proteins remains to be elucidated and it has been suggested that there is a large degree of functional redundancy between SPI2 effector proteins (Haraga et al., 2008); however it is clear that SPI2 is essential during Salmonella proliferation within all cell types, to resist the antimicrobial mechanisms of the host immune system (Hautefort et al., 2008, Haraga et al., 2008). A key function of SPI2-secreted effector proteins is to modulate SCV maturation and to alter host cell endocytic trafficking. The endocytic pathway, which ends with lysosomal fusion, normally functions to degrade intracellular vescicles. The association of some endocytic markers with the SCV indicates that the SCV selectively interacts with parts of the endocytic pathway but evades destruction by lysosomal fusion (Abrahams & Hensel, 2006). The SPI2 effector SsaB (SpiC), which also functions as part of the SPI2 apparatus, is translocated into the host cytosol where it inactivates the mammalian protein Hook3, resulting in a blocking of the fusion of the SCV and lysosome (Shotland *et al.*, 2003).

Translocation of effector proteins via the SPI2 TTSS results in host actin cytoskeleton rearrangements that lead to the formation of an actin meshwork around the SCV. This process is known as vacuole-associated actin polymerisation (VAP) and is required for maintaining the stability and integrity of the SCV membrane. Effectors encoded outside SPI2, SspH2 and SseI, as well as the plasmid-encoded SpvB and the SPI1-secreted actin binding protein SipA play a role in VAP (Fabrega & Vila, 2013). During maturation, the SCV makes use of the host microtubule-associated proteins kinesin and dynein to migrate to a juxtanuclear position, where bacterial replication takes place, likely due to the increased access to nutrients and membrane fragments (Haraga et al., 2008). The SPI2 effectors SseG and SseF play a role in maintaining the SCV near the nucleus and Golgi apparatus, while the SifA protein contributes by limiting the amount of kinesin recruited to the SCV site, thus avoiding displacement of the SCV to the cell periphery (Abrahams & Hensel, 2006). The microtubule cytoskeleton also acts as a scaffold for the formation of SIFs, which are an elongation of the SCV and are derived from endocytic compartment membranes (Krieger et al., 2014). SifA localises along the microtubules and maintains the integrity of the SIFs (Brumell et al., 2002). The ability of SPI2-secreted effector proteins to modulate the motility of infected macrophages plays an important role in bacterial dissemination and the establishment of systemic infection within the liver and spleen (Fabrega & Vila, 2013).

#### 1.2 Transcriptional Regulation

#### 1.2.1 Methods of transcriptional regulation

Transcription initiation is considered to be one of the most important control points for gene expression. The co-ordinated regulation of transcription initiation is an essential mechanism to optimise the use of valuable nutritional resources, to adapt to rapidly changing environments and to avoid unnecessary energy consumption (Seshasayee *et al.*, 2011). As previously discussed, the infectious process of *S.* Typhimurium involves a series of stressful events for the bacterium. *S.* Typhimurium encounters an array of different environmental conditions from survival within animal intestinal tracts, to

survival in soil, water or on contaminated food, to ingestion by the human host. Upon ingestion the bacterium experiences temperature and pH shock, competition with host microbiota, antimicrobial peptides, nutrient starvation, ROS, RNS and host phagocytic cells (Fabrega & Vila, 2013). The ability of a bacterial pathogen to survive such harsh host environments and to establish a successful infection depends not only on the virulence factors that the pathogen expresses, but also on the ability of the pathogen to control when and where these genes are expressed (Groisman & Mouslim, 2006). Genes involved in survival and resistance to stressful environmental conditions and virulence-associated genes must be expressed at the correct time and in the correct location to avoid causing a fitness defect for the bacterium, and *Salmonella* must carefully co-ordinate its programme of gene expression in order to survive each harsh condition and to establish a successful infection. Adapting to new environments can require promoting the expression of some genes to aid in bacterial survival, while repressing the expression of other genes that are helpful under certain conditions, but may be detrimental under others (Groisman & Mouslim, 2006).

An example of the importance of ensuring the correct temporal and spatial conditions for gene expression results from the phenotypic variation or bistability of SPI1 gene expression (Hautefort *et al.*, 2003), in which only a portion of genetically identical *Salmonella* cells express the genes necessary for host epithelial cell invasion and induction of intestinal inflammation. It was later shown that the proportion of cells that do express SPI1 are at a fitness-disadvantage, compared to those that do not express SPI1 genes. The fitness defect occurs as a result of the expression of genes encoding the translocon and effector proteins (Sturm *et al.*, 2011). It was suggested that the bistability and fitness cost imposed by expression of the SPI1 TTSS may be a mechanism to allow the cells which do not express the TTSS to benefit from the induced pro-inflammatory response and enhanced bacterial dissemination. These experiments were performed under laboratory conditions that should induce expression of SPI1 but the data highlight the burden of expressing virulence genes, even under conditions when expression of these genes is necessary, and suggest that a similar fitness cost would be imposed by expression of virulence genes under non-infection relevant conditions.

There are countless mechanisms that regulate the genetic output of bacterial cells, and the key mechanisms of gene regulation used by the regulators that are relevant for this study

are outlined below and illustrated in Figure 1.2. Transcriptional activation is mediated by DNA-dependent RNA polymerase (RNAP), and many methods of regulating the activation or repression of transcription are based on the properties of RNAP and/or the ability of a protein to target and recruit RNAP. Initiation of transcription requires the RNAP holoenzyme, which consists of the RNAP apoenzyme ( $\beta\beta\alpha_2\omega$ ) and the  $\sigma$ -subunit which is necessary for promoter recognition, DNA binding and open complex formation (Lee et al., 2012). The RNAP holoenzyme is necessary and sufficient to initiate transcription in the presence of an optimal promoter sequence and structure, which we previously defined for S. Typhimurium  $\sigma^{70}$ -driven promoters (Kröger et al., 2012). Deviations from the optimal consensus promoter leads to a requirement for additional layers of regulation, allowing promoters to detect the correct temporal and spatial signals for gene expression, as gene promoters are often responsive to specific environmental cues or are regulated by a protein that is expressed in response to specific environmental cues (Seshasayee et al., 2011).

Alternating the DNA-binding  $\sigma$ -subunit of the RNAP holoenzyme provides one of the most basic methods of selectively regulating a subset of genes in response to changes in environmental conditions (Lee et al., 2012). In Salmonella there are 2 families of  $\sigma$ -factor:  $\sigma^{70}$  and  $\sigma^{54}$ . The  $\sigma^{70}$  family comprises the majority of Salmonella  $\sigma$ -factors, including the house-keeping  $\sigma$ -factor RpoD ( $\sigma^{70/D}$ ) and the alternative  $\sigma$ -factors RpoS  $(\sigma^{38/S})$ , RpoE  $(\sigma^{24/E})$ , RpoH  $(\sigma^{32/H})$ , and FliA  $(\sigma^{28})$  (Bang et al., 2005), while the  $\sigma^{54}$ family contains only one alternative σ-factor (RpoN) (Seshasayee et al., 2011). Most promoters contain motifs optimal for recognition and binding by  $\sigma^{70}$ . Alternative  $\sigma$ -factors, such as  $\sigma^{38}$  and  $\sigma^{24}$ , can re-programme the preference of the RNAP holoenzyme towards particular promoter motifs or structures (Lee et al., 2012). The  $\sigma$ -factors compete for limiting amounts of the RNAP apoenzyme molecule. The outcome of the competition is determined by various factors, such as availability or activity of the  $\sigma$ -factor or anti- $\sigma$ -factor, and the tolerance of the  $\sigma$ -factor for promoter sequences which are divergent from the consensus sequence (Seshasayee et al., 2011). Promoters which recognise  $\sigma^{54}$ -associated RNAP differ from those which recognise  $\sigma$ -factors of the  $\sigma^{70}$ family (Lee et al., 2012). These differences are illustrated in Figure 1.2 B and will be further discussed in section 1.2.7.

Another method of selectively regulating transcription initiation is the use of *trans*-acting transcription factors (TF). TFs, typically, are DNA-binding proteins that recognise and bind specific DNA sequences in the regulatory regions of target genes and modulate transcription of those target genes. Activation of the TF, and subsequent regulation of the target genes of that TF, generally occurs as a result of an extracellular or intracellular trigger (Lee et al., 2012). TFs can act on a global scale, targeting large numbers of genes, including genes encoding other regulatory proteins, or TFs can act on a local scale, targeting only a small number of genes in response to a specific cue (Seshasayee et al., 2011). The DNA-Binding Domain Database (DBD) predicts that approximately 6% (290 genes) of the *S.* Typhimurium LT2 genome encodes validated or hypothetical TFs (Wilson *et al.*, 2008). The number of TFs decreases in prokaryotes that are associated with a parasitic or symbiotic lifestyle, as environmental conditions are unlikely to change as rapidly and dramatically for parasitic or symbiotic organisms (Seshasayee et al., 2011) while *S.* Typhimurium regularly encounters rapidly changing environmental conditions.

TFs can modulate transcription by acting as an activator or as a repressor. Some TFs act solely as activators or repressors of gene expression, while other TFs can act either as an activator or repressor depending on the target gene promoter (Browning & Busby, 2004). TFs directly control transcription initiation through interaction with subunits of RNAP. Activators typically act by binding to sites which overlap the -35 hexamer within the target gene promoter and interact with domain 4 of the  $\sigma$ -subunit of RNAP. Activating TFs may also bind to the  $\alpha$ -subunit C-terminal domain ( $\alpha$ -CTD) which is connected to the rest of the RNAP complex by a non-structured flexible linker region (Seshasayee et al., 2011). Association of a TF with the  $\alpha$ -CTD provides flexibility in the location of the TF binding site (Lee et al., 2012). In *E. coli*, FNR and PhoB are associated with the  $\sigma$ -subunit (Makino *et al.*, 1993, Lonetto *et al.*, 1998), while some AraC-like proteins, such as HilD in *S.* Typhimurium, can target the  $\alpha$ -CTD (Olekhnovich & Kadner, 2004). OmpR requires the  $\alpha$ -CTD of RNAP to activate transcription (Kenney, 2002).

Some methods of regulation of transcription can be indirect, for example the methylation state of DNA can act as a signal for RNAP to initiate transcription (Low & Casadesus, 2008), while Dam-mediated DNA methylation also functions as an anti-repressor of H-NS at the promoter of the plasmid-encoded antisense RNA FinP (Camacho *et al.*, 2005). In fact, the most common indirect method of initiating transcription is to relieve

repression. Nucleoid-associated proteins (NAPs), such as H-NS, mediate transcriptional silencing by binding DNA and forming structures which inhibit the progress of RNAP along a gene. H-NS-mediated formation of inhibitory structures is a common method of transcriptionally silencing horizontally-acquired DNA to avoid inappropriate expression of foreign DNA which could disrupt the established cellular networks and cause a reduction in fitness of the recipient bacteria (Lucchini et al., 2006, Navarre et al., 2007). However, expression of horizontally-acquired sections of DNA can be beneficial to the bacteria under certain circumstances, and so mechanisms of relieving the H-NS-mediated transcriptional silencing have evolved (Stoebel et al., 2008). Counter-silencing can be mediated in a protein-independent fashion due to environmental stimuli such as high temperatures or osmolarity, which directly affect DNA structure (Stoebel et al., 2008). However, protein-dependent mechanisms of counter-silencing are also common. The SlyA family of proteins acts to counter-silence H-NS by different mechanisms in a number of bacterial species. SlyA displaces H-NS from the hlyE promoter in E. coli (Lithgow et al., 2007) and re-structures the H-NS-bound promoters pagC and ugtL in S. Typhimurium (Perez et al., 2008), allowing RNAP to proceed with transcription initiation. SlyA does not appear to have the ability to activate transcription by itself and, therefore, often requires binding by one or more other proteins to activate transcription after relief of the H-NS-mediated silencing (Stoebel et al., 2008). A subset of PhoP-dependent genes requires SlyA-mediated counter-silencing, while PhoP associates with the  $\alpha$ -CTD of the RNAP holoenzyme to activate transcription (Zwir et al., 2012). The co-regulation by PhoP and SlyA is in contrast to SsrB which has the dual function of counter-silencing H-NS and activation of transcription by direct interaction with different subunits of RNAP (Walthers et al., 2007).

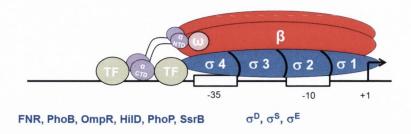
A variation on the theme of counter-silencing involves inhibiting the activities of a repressor protein. FliZ activation of class 2 flagellar operons occurs via this method. FliZ binds and represses transcription from the ydiV promoter. The ydiV gene encodes an anti-FlhD<sub>4</sub>C<sub>2</sub> factor which binds to the FlhD subunit of the master flagellar regulator, resulting in inhibition of transcription of class 2 flagellar operon genes. FliZ binds a sequence which overlaps the -35 site of the ydiV promoter, presumably occluding RNAP from the site. The prevention of ydiV transcription results in an increase in expression of class 2 flagellar operons (Wada  $et\ al.$ , 2011). Environmental conditions also affect the activity of repressor proteins. Fur acts as a transcriptional repressor when levels of the

ferrous ion are high. Fur dimerises when associated with the ferrous ion, and binds a "Fur box" within target promoters, blocking the binding of RNAP and thus preventing transcription of these genes. When iron levels are low, iron dissociates from Fur and Fur is displaced from the DNA, permitting RNAP binding and transcription initiation (Carpenter *et al.*, 2009).

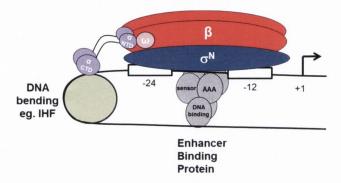
Recently, a novel mechanism which affects transcriptional regulatory control has been described in E. coli, and also occurs, albeit at a lower frequency, in S. enterica. This mechanism is known as horizontal regulatory transfer (HRT) and involves the acquisition of divergent non-homologous promoter regions through lateral DNA transfer (Oren et al., 2014). Horizontal or lateral gene transfer has traditionally been thought of as the key to microbial evolution and diversity (Navarre et al., 2007) and will be further discussed in section 5.1.1. Horizontal gene transfer involves the transfer of coding sequences, and has provided S. enterica with key virulence genes and pathogenicity islands leading to the formation of the S. enterica pan-genome (Jacobsen et al., 2011). HRT, by contrast, involves the transfer of transcriptionally and translationally silent regulatory regions and is ubiquitous in promoters of ancestral genes, challenging the assumptions that ancestral genes do not contribute to bacterial diversification. The regulatory switching mechanism is thought to be mediated by homologous recombination, as the flanking genes are typically highly conserved, even across distant species (Oren et al., 2014). HRT appears to serve as a mechanism of altering promoter architecture, promoting alternative TF binding or altering TF binding specificity, with the effect of providing fitness advantages to recipient strains in certain niches, thus contributing to diversification of strains within a species as well as interspecies variation (Oren et al., 2014).

The previously described examples of mechanisms of transcriptional regulation represent a subset of the methods that are used by S. Typhimurium to regulate gene expression at the transcriptional level. The multiple layers of regulation that can occur at a single gene promoter highlight the complexity of cellular regulatory networks. This study involves the investigation of the regulons of 16 important virulence-associated S. Typhimurium  $\sigma$ -factors and TFs using an RNA-seq-based transcriptomics approach. The following section will review the roles of these 16  $\sigma$ -factors and TFs, in terms of virulence, and is not intended to be an exhaustive review of all S. Typhimurium TFs. The main regulatory interactions concerning proteins relevant for this study are shown in Figure 1.3.

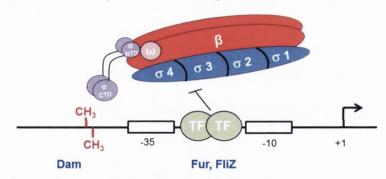
**A.** Activating TFs interact with the  $\alpha$ -CTD or with domain 4 of the  $\sigma$ -subunit of the RNA polymerase (RNAP) holoenzyme.



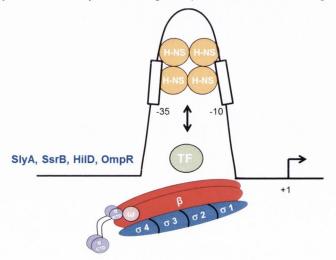
**B.**  $\sigma^{N}$ -dependent promoters require enhancer binding proteins (EBP). DNA bending proteins facilitate contact between the EBP and the  $\sigma$ -subunit of the RNAP holoenzyme.



**C.** Repressing TFs prevent the RNAP holoenzyme from binding promoter regions. Dam-mediated DNA methylation acts as a signal for RNAP to repress or initiate transcription.



**D.** NAPs form inhibitory nucleo-protein structures that prevent transcription initiation by RNAP, but may be relieved by TF binding and promoter re-structuring or NAP displacement.



# Figure 1.2 Key mechanisms involved in regulating the initiation of bacterial transcription

**A.** The RNA polymerase holoenzyme consisting of  $\sigma^D$ ,  $\sigma^S$ , or  $\sigma^E$  subunits binding to the -10 and -35 hexamers of a gene promoter. Activating transcription factors (examples listed in blue) can make contact with the  $\sigma$  or  $\alpha$  subunits of RNAP to activate transcription. The  $\sigma$ -subunit is divided into 4 domains, TFs typically make contact with domain 4 (Lee *et al.*, 2012). **B.** RNAP containing a  $\sigma^{54}$  subunit recognises hexamers at -12 and -24 sites in gene promoters and requires binding by an enhancer binding protein (EBP) to activate transcription. DNA bending proteins may be used to bring the EBPs in contact with the  $\sigma^{54}$  subunit. **C.** Repressor proteins (examples listed in blue) can occlude promoters and prevent RNAP binding. Dam-mediated DNA methylation (CH<sub>3</sub>) may act as a mark to repress or activate gene transcription. **D.** Nucleoid-associated proteins (NAPs), such as H-NS, can form inhibitory complexes in promoter regions. Counter-silencing proteins (examples listed in blue) can displace H-NS or re-structure the promoter to allow RNAP to gain access and activate transcription. Promoter recognition sites (-10 and -35 or -12 and 24) are indicated upstream of the transcription start site (TSS). Bent arrows and +1 denote the TSS.

#### 1.2.2 Dam

The DNA adenine methyltransferase protein, Dam, catalyses the methylation of adenine residues of DNA, using S-adenosyl-L-methionine as a methyl donor (Marinus & Casadesus, 2009). Dam typically targets the N6 position of the adenine residue of GATC sites of hemi-methylated DNA, where one strand is methylated and the other is not, as occurs in post-replicative DNA (Low & Casadesus, 2008). Adenine methylation reduces the thermodynamic stability of DNA and alters DNA curvature, affecting DNA-protein interactions at or near these sites (Marinus & Casadesus, 2009). RNAP and other TFs can also recognise the methylation state of DNA and this can be used to couple gene regulation to certain phases of the cell cycle (Low & Casadesus, 2008). Dam-mediated discrimination between fully methylated and hemi-methylated DNA is important for the post-replicative mismatch repair machinery to identify and repair errors in the newly synthesised DNA strand (Marinus & Casadesus, 2009). Adenine methylation is important during the chromosome replication process. The region of DNA comprising the origin of replication (oriC) and the gene encoding DnaA, the main activator of DNA replication, contains a relatively large number of GATC sites which must be fully methylated for the initiation of replication to occur. A second hemi-methylated

DNA-specific protein, SeqA, sequesters the hemi-methylated DNA binding sites after replication to prevent premature firing from the origin of replication during each cell cycle. Sequestration by SeqA prevents Dam methylation and sequestration must be overcome to allow Dam to fully methylate the DNA and initiate a new round of chromosome replication (Low & Casadesus, 2008).

Dam-mediated methylation and the DNA methylation state affects DNA transactions, such as transposition, phase variation of pilus-encoding genes in uropathogenic  $E.\ coli$  and protection from transcriptional silencing by the NAP H-NS (Marinus & Casadesus, 2009). Dam plays a role in S. Typhimurium virulence as absence of the Dam protein results in severe attenuation in the mouse model via the oral and i.p. infection routes due to impaired epithelial cell invasion (Garcia-Del Portillo  $et\ al.$ , 1999). Reduced levels of the primary SPI1 regulator HilD, and a subsequent reduction in expression of the SPI1 TTSS, due to post-transcriptional regulation by Dam are likely to be the cause of the reduced invasion phenotype in a  $\Delta dam$  background (Balbontin  $et\ al.$ , 2006, Lopez-Garrido & Casadesus, 2010). Absence of Dam methylation in S. Typhimurium also results in a reduced motility phenotype (Balbontin et al., 2006), over-expression of the std fimbrial operon (Jakomin  $et\ al.$ , 2008) and reduced resistance to bile (Prieto  $et\ al.$ , 2004).

In this study the role of Dam methylation was investigated during logarithmic growth in rich media  $(OD_{600}0.3;\ MEP)$  as this growth phase maximises the number of hemi-methylated GATC sites, as a result of active replication.

#### 1.2.3 HilD

HilD is an AraC-like protein encoded on SPI1 (Ellermeier & Slauch, 2007). The current model of HilD-mediated gene regulation posits that HilD forms a feed-forward regulatory loop with 2 other AraC-like regulators, the SPI1-encoded HilC and RtsA, which is encoded elsewhere on the chromosome (Ellermeier *et al.*, 2005). The HilD, HilC and RtsA proteins can each activate expression of *hilD*, *hilC* and *rtsA*. The regulatory feed-forward loop also activates expression of the central SPI1 regulator, HilA. HilA activates all necessary components of a functional SPI1 TTSS directly, through activation of the *prg/org* apparatus-encoding genes, and indirectly, through activation of InvF which subsequently activates the SPI1-located *sic/sip* effector-encoding operon, as well as

effectors encoded elsewhere on the chromosome (Ellermeier & Slauch, 2007). HilD, HilC and RtsA bind the hilA promoter region (Olekhnovich & Kadner, 2002, Ellermeier & Slauch, 2003). It has been suggested that HilD and HilC activate hilA expression by over-coming silencing by the NAPs H-NS and Hha (Olekhnovich & Kadner, 2006, Schechter & Lee, 2001). This view has been contradicted, however, as hilA expression requires HilD in the absence of negatively-regulating proteins, and HilD interacts with the α-CTD of RNAP at the hilA promoter. Thus HilD has the dual role of activating hilA expression by a counter-silencing mechanism and by recruitment or positioning of RNAP at the hilA promoter (Boddicker et al., 2003, Olekhnovich & Kadner, 2004). feed-forward loop also has a HilA-independent role in virulence gene regulation, as HilD, HilC and RtsA can directly activate expression of the SPI1-encoded inv/spa operon and the genes encoding the SPI1-translocated effector SlrP and the disulphide oxidoreductase DsbA (Ellermeier & Slauch, 2007). Thus, the feed-forward loop comprising HilD, HilC and RtsA is necessary to induce SPI1 gene expression and to activate SPI1 secreted proteins, allowing S. Typhimurium to invade host intestinal epithelial cells and induce intestinal inflammation. HilD also mediates cross-talk between SPI1 and SPI2 via counter-silencing of H-NS at the promoter of the gene encoding the major SPI2 regulator, SsrB (Bustamante et al., 2008, Martinez et al., 2014).

Furthermore, HilD activates expression of 2 small non-coding regulatory RNAs (sRNA), InvR and DapZ. InvR is SPI1-encoded but post-transcriptionally regulates expression of the ancestral outer membrane protein (OMP) gene, *ompD*, while DapZ is encoded outside of SPI1 and regulates expression of the genes encoding ancestral ABC transporters (Pfeiffer *et al.*, 2007, Chao et al., 2012). These studies indicate that, in addition to the function of HilD as a major virulence determinant of *S*. Typhimurium, HilD can co-opt the function of genes encoded on the ancestral *S*. Typhimurium chromosome, presumably to modulate survival of *S*. Typhimurium during the infection process. HilD-mediated regulation of genes encoded outside of SPI1 was recently demonstrated by a ChIP-seq-based study that found the majority of HilD binding sites are located outside of the SPI1 island (Petrone *et al.*, 2014).

The role of HilD is predominant over the roles on HilC and RtsA in the complex SPI1 feed-forward regulatory loop, as the absence of the HilD protein results in a greater reduction in *hilA* expression than the absence of HilC or RtsA (Ellermeier et al., 2005).

As HilD acts at the top of the SPI1 regulatory hierarchy, it was chosen for investigation in this study. Signals that result in condition-specific induction of the SPI1 TTSS are integrated at the level of HilD regulation (Ellermeier & Slauch, 2007). SPI1 gene expression is activated in the host intestine or under laboratory conditions which mimic the host intestinal environment. High osmolarity and low oxygen are typically considered as "SPI1-inducing" conditions (Bajaj *et al.*, 1996) and these conditions are achieved through growth in rich media to an  $OD_{600}2.0$  (ESP). As a result of the reported HilD-mediated cross-talk between SPI1 and SPI2, the HilD regulon under SPI2-inducing conditions (growth to  $OD_{600}0.3$  in minimal PCN media (Lober et al., 2006)) was also investigated in this study.

#### 1.2.4 FliZ

FliZ is a transcriptional regulator involved in control of flagellar gene expression and is encoded within an operon that also encodes the flagellum-specific alternative  $\sigma$ -factor, FliA (Ikebe et al., 1999). The genes involved in flagellar gene expression are divided into 3 classes according to the hierarchy of transcription: early, middle and late and the corresponding promoters are classified as class1, 2 or 3 respectively (Chilcott & Hughes, 2000). fliZ belongs to the middle class of flagellar genes and is transcribed from 2 promoters. fliAZY is transcribed from a class 2 gene promoter which is directly activated by  $\sigma^{70}$  and by the early class master flagellar regulators, FlhDC. There is also a class 3 promoter which is  $\sigma^{28}$  (FliA)-dependent (Ikebe et al., 1999). FliZ activates expression of middle class genes from class 2 promoters, as well as, activating expression of FlhC at the post-transcriptional level, most likely through protein stabilisation (Saini et al., 2008). Expression of fliZ from a class 3 promoter and the regulation of FlhDC protein levels by FliZ provides regulatory loops that allow the assembly of flagella to be co-ordinated to ensure that cells express a sufficient amount of flagella and to ensure that progeny cells can rapidly assemble flagella. Activation of class 2 promoters by  $\sigma^{28}$  is mediated through activation of fliZ (Saini et al., 2008).

As previously discussed, flagella play a role in *S*. Typhimurium virulence as motility is necessary for bacterial approach to the epithelial cells of the host ileum, which is a prerequisite for bacterial adhesion and invasion of host cells. FliZ, therefore, is an

important virulence determinant of S. Typhimurium due to the FliZ-mediated activation of flagellar assembly. In addition to the role of FliZ in flagellar assembly, FliZ also contributes to S. Typhimurium virulence by activation of SPI1 via control of HilD expression (Kage  $et\ al.$ , 2008). The activation of SPI1 by FliZ is both post-transcriptional and post-translational. It is thought that FliZ somehow affects the activity of the HilD protein and FliZ also has a mild effect on stability of the HilD protein. Absence of the FliZ protein results in attenuation during oral infection of the mouse, but this effect is only partially dependent on SPI1. The  $\Delta fliZ$  mutant and wild-type compete equally well following i.p. infection of mice, highlighting that FliZ plays a role in intestinal colonisation and/or invasion of host epithelial cells but FliZ does not contribute to intracellular survival or the systemic phase of S. Typhimurium infection (Chubiz  $et\ al.$ , 2010).

In order to investigate the role of FliZ in terms of its control of SPI1 and invasion-associated virulence, the  $\Delta fliZ$  mutant in this study was grown under SPI1-inducing conditions (ESP) and FliZ is referred to as a SPI1-associated regulator.

#### 1.2.5 BarA/SirA

BarA is the sensor kinase and SirA is the response regulator of an ancestral two-component system (TCS) involved in the activation of SPI1 gene expression (Ahmer et al., 1999, Altier et al., 2000b, Johnston et al., 1996). The current model for SPI1 gene activation by the BarA/SirA TCS is that SirA directly binds to and presumably regulates expression from the hilA and hilC promoters (Teplitski et al., 2003) but that SirA also indirectly activates invasion gene expression through activation of the sRNAs CsrB and CsrC (Fortune et al., 2006). The RNA binding protein CsrA modulates SPI1 expression through binding hilD mRNA in the translation initiation region, preventing ribosome binding and translation initiation and leading to increased turnover of hilD mRNA (Martinez et al., 2011). However, the method of regulation of HilD by CsrA appears to be complex, as both the absence and over-expression of CsrA results in a decrease in HilD expression, suggesting that regulation of CsrA expression must be tightly controlled for optimal invasion gene expression (Altier et al., 2000a). CsrB and CsrC directly and redundantly bind the CsrA protein and titrate CsrA from its targets to prevent

CsrA-mediated repression of invasion gene expression. The predominant method of SPI1 activation by BarA/SirA is via control of the *csr* system as opposed to the direct activation of invasion gene expression by SirA (Fortune et al., 2006). The extracellular signals which induce BarA phosphorylation are, as yet, unknown but the short chain fatty acid acetate, which is present at high concentrations at the preferred site for *S*. Typhimurium infection, the ileum, activates SirA expression (Lawhon *et al.*, 2002). This activating mechanism occurs through the formation of acetyl phosphate from acetate, resulting in the phosphorylation of SirA. The acetate-mediated activation of SirA expression appears to be independent of BarA, but suggests that SirA integrates one or more appropriate environmental signals under conditions when expression of invasion genes is necessary (Lawhon et al., 2002).

The BarA/SirA TCS is, therefore, considered to be a SPI1-associated regulatory system, although BarA and SirA are not encoded on SPI1. In this study the  $\Delta barA/sirA$  mutant was grown under SPI1-inducing conditions (ESP) to further elucidate the SPI1-associated function of this regulatory system.

#### 1.2.6 Fur

The ferric uptake regulator, Fur, binds divalent cations, typically the ferrous ion (Fe<sup>2+</sup>). When iron levels are high, Fe<sup>2+</sup>-bound Fur binds to the promoter regions of target genes, in sites that overlap the RNAP binding site, resulting in repression of transcription of the target genes. Fur targets genes involved in siderophore production, iron transport and iron acquisition, thus under iron-replete conditions additional iron is not imported into the cell. When iron levels are low, Fur dissociates from the ferrous ion and also from its target DNA, allowing transcription of the genes involved in acquisition of iron. In this way, Fur achieves iron homeostasis within the cell, maintaining levels of iron which acts as a crucial co-factor in many cellular processes, but decreasing the risk of iron toxicity (Kadner, 2005). In *E. coli* Fur represses expression of an sRNA, RyhB, which represses a number of genes encoding iron storage and iron usage proteins. These genes are positively regulated by Fur via an indirect mechanism. In iron-replete conditions, Fur is bound to and represses expression from the *ryhB* promoter, resulting in de-repression/expression of the RyhB-controlled target genes. When iron levels are

reduced Fur-mediated repression of RyhB is lifted and the genes involved in iron storage and iron usage are repressed by RyhB (Masse & Gottesman, 2002, Masse *et al.*, 2005). In *S.* Typhimurium RyhB-1 (RfrA) and RyhB-2 (RfrB) are the *E. coli* RyhB orthologue and paralogue. Similar to *E. coli* RyhB, S. Typhimurium RyhB-1 and RyhB-2 repress translation of the superoxide dismutase-encoding *sodB* transcript (Ellermeier & Slauch, 2008).

Iron is an essential element for all organisms and many pathogens preferentially colonise niches with accessible iron stores, such as human hosts. The important role played by iron in virulence and host-pathogen interactions is exemplified by the strong association between host iron overload and host susceptibility to bacterial infections (Carpenter et al., Fur plays an important role in S. Typhimurium virulence by activating 2009). transcription of hilA via the HilD protein (Ellermeier & Slauch, 2008). Fur represses transcription of the hns gene, and it has been suggested that the reduction in hilA transcription in the absence of the Fur protein is due to a de-repression of hns H-NS silences expression of hilA and HilD plays a role by transcription. counter-silencing H-NS at the hilA promoter (Troxell et al., 2011b). However, there are conflicting reports that metal-bound Fur controls SPI1 expression through direct binding to the hilD promoter and activation of hilD transcription (Teixido et al., 2011), although S. Typhimurium Fur has never been previously shown to act as a direct transcriptional In addition, the original study which demonstrated a HilD-mediated link between Fur and SPI1 reported no effect on hilD transcription in the absence of the Fur protein (Ellermeier & Slauch, 2008).

A  $\Delta fur$  mutant is attenuated for virulence following i.p. infection of mice, suggesting that Fur may play a further role in S. Typhimurium virulence in addition to controlling SPII gene expression (Velayudhan et~al., 2007). It is tempting to speculate that the reduction in virulence in the absence of the Fur protein may reflect the de-repression of the H-NS protein, leading to a general silencing of pathogenicity island-encoded genes (Troxell et al., 2011b). However, Fur does repress expression of SPI2 genes via direct binding of the ssrB promoter under acidic conditions. Absence of the Fur protein allowed SPI2 genes to become induced earlier and to a higher level within macrophages (Choi et~al., 2014). In addition to controlling the transcription of genes which encode important virulence factors, Fur also regulates expression of a number of genes that are important for

intracellular survival such as genes encoding components of the ATR and resistance to oxidative stress (Carpenter et al., 2009).

We wished to further investigate the important role of Fur-mediated SPI1 activation, thus the  $\Delta fur$  mutant was grown under SPI1-inducing conditions (ESP) and is considered to be a SPI1-associated regulator for the purposes of this study.

## $1.2.7 \, \sigma^{54}$

 $\sigma^{54}$  (RpoN) is structurally and functionally distinct from members of the  $\sigma^{70}$  family of sigma factors, and genes encoding  $\sigma^{54}$  are only found in approximately 60% of sequenced bacterial genomes (Merrick, 1993, Lee et al., 2012).  $\sigma^{54}$  is necessary for transcription of a number of genes involved in a variety of cellular functions including nitrogen assimilation and fixation (Morett & Segovia, 1993). The  $\sigma^{70}$  family of  $\sigma$ -factors recognise and bind hexamers located -10 and -35 nucleotides (nt) upstream of the transcriptional start site (TSS), however,  $\sigma^{54}$ -bound RNAP recognises and binds sites centered at -12 and -24 nt upstream of the TSS. The -12 and -24 hexamers are too close to each other, and bound RNAP is unable to generate a stable open complex when bound at these sites (Lee et al., 2012). Thus isomerisation to a transcriptionally-competent open complex requires an enhancer binding protein (EBP) bound to a site typically located between 100 and 200 base pairs (bp) upstream of the TSS, but these sites can be functional at even longer distances (Morett & Segovia, 1993). EBPs are modular proteins that consist of three domains: the C-terminal domain contains a helix-turn-helix (H-T-H) DNA binding motif, a conserved central domain that contains an AAA+ activator domain and the N-terminal domain that is responsible for detecting metabolic and other regulatory signals (Lee et al., 2012). The EBPs assemble into hexamers and bind the upstream regulatory region of the DNA. The activator domain of the EBP must make contact with the RNAP-bound  $\sigma^{54}$  so DNA bending proteins, such as IHF (Integration Host Factor) are often required to generate curvature in the DNA to bring the EBP and  $\sigma^{54}$ -subunit into contact. ATP hydrolysis mediated by the AAA+ activator domain of the EBP is necessary during the transition from closed to open complex and transcription initiation (Figure 1.2 B) (Lee et al., 2012).

In a number of members of the enterobacteriaceae family, including S. Typhimurium, RpoN controls expression of 2 sRNAs, GlmY and GlmZ, that are involved in activation of the glmS gene which encodes glucosamine 6-phosphate synthase (Gopel et~al., 2011).  $\sigma^{54}$  is associated with virulence gene regulation in many bacterial species, for example  $\sigma^{54}$  is an essential colonisation factor and activates expression of motility-associated genes and the virulence-associated type 6 secretion system in the pathogen Vibrio~cholerae (Dong & Mekalanos, 2012). Additionally, subsets of flagellar genes are transcribed from  $\sigma^{54}$ -dependent promoters in a number of bacterial pathogens, including Helicobacter~pylori, Campylobacter~jejuni and Pseudomonas species (Tsang & Hoover, 2014).  $\sigma^{54}$  regulates expression of the rfaH gene in S. Typhi, which is necessary for the production of O-antigen lipopolysaccharide (LPS), a major component of the bacterial cell envelope, and important for epithelial cell invasion (Bittner et~al., 2002). In order to investigate if S. Typhimurium  $\sigma^{54}$  plays a role in controlling expression of genes under invasion-associated conditions, the  $\Delta rpoN$  mutant was grown to ESP in this study.

## 1.2.8 $\sigma^{38}$

 $\sigma^{38}$  (RpoS) is a member of the  $\sigma^{70}$  family of  $\sigma$ -factors and recognises and binds promoter elements centered around -10 and -35 nt relative to the TSS (Lee et al., 2012). There is very little variation in the promoter consensus sequences recognised by  $\sigma^{38}$  and  $\sigma^{70}$ , however  $\sigma^{38}$  has greater tolerance for deviations in sequence and structure from the consensus sequence and structure, allowing for selective transcription by either  $\sigma$ -factor under appropriate conditions (Typas *et al.*, 2007).  $\sigma^{38}$  becomes active in stationary phase and in response to a variety of other stressful conditions, and mediates transcription of genes involved in the general stress response (Hengge-Aronis, 2002). Expression of the *rpoS* gene is strictly controlled at the level of transcription and to a greater extent at the post-transcriptional level via a complex network of signals and regulatory factors. In *E. coli*, cyclic AMP (cAMP), polyphosphate and guanosine 3',5'-bispyrophosphate (ppGpp) are some of the factors thought to affect *rpoS* transcription (Hengge-Aronis, 2002). At the post-transcriptional level in *E. coli*, the sRNAs DsrA, RprA and ArcZ activate *rpoS* translation by relieving an inhibitory structure masking the ribosome binding site (RBS) (Majdalani *et al.*, 1998, Majdalani *et al.*, 2002, Mandin & Gottesman, 2010), while OxyS

inhibits rpoS expression (Altuvia et~al., 1997). However, the role of DsrA and RprA are not as important for activation of rpoS expression in S. Typhimurium (Jones et~al., 2006). The stability of  $\sigma^{38}$  is also regulated by ClpXP protease-mediated proteolysis (Hengge-Aronis, 2002).

In addition to mediating transcription of genes involved in survival of stressful environmental conditions,  $\sigma^{38}$  also regulates virulence-associated genes in S. Typhimurium, perhaps highlighting the stresses encountered by pathogenic bacteria during infection. The control of stress response genes, such as resistance to oxidative stress, low pH, DNA damage and starvation, by  $\sigma^{38}$  mediates S. Typhimurium survival in the host intracellular environment (Dong & Schellhorn, 2010).  $\sigma^{38}$  also targets genes that encode specific virulence factors. The spvABCD locus is encoded on the Salmonella pSLT virulence plasmid and the gene products, SpvB and SpvC, are translocated via the SPI2 TTSS. The spv locus is required for intracellular survival, persistence and NTS bacteraemia (Guiney & Fierer, 2011). The spvABCD operon is transcribed from a  $\sigma^{38}$ -dependent promoter and requires the TF SpvR, which is also transcribed from a  $\sigma^{38}$ -dependent promoter (Fang et al., 1992, Chen et al., 1995). Additionally,  $\sigma^{38}$ positively regulates the transcription from the csgDEFG and csgBA promoters, which encode components necessary to make curli fibres for host cell adhesion (Romling et al., 1998). A  $\Delta rpoS$  mutant is attenuated following oral infection of mice, partly as a result of reduction in spv gene expression, as well as the expression chromosomally-encoded  $\sigma^{38}$  target genes (Fang et al., 1992, Kowarz et al., 1994).

In this study the  $\sigma^{38}$  regulon was investigated in late stationary phase (LSP), which entails growth in rich media to an  $OD_{600}2.0$ , followed by a further 6 hours of growth. LSP is the condition when  $\sigma^{38}$  expression is highest due to a reduction in proteolytic degradation (Hengge-Aronis, 2002), thus we would also expect many of the genes within the  $\sigma^{38}$  regulon to also be induced under these conditions.

# $1.2.9 \sigma^{24}$

 $\sigma^{24}$  (RpoE) is another member of the  $\sigma^{70}$  family of  $\sigma$ -factors. RpoE is known as the extracytoplasmic function alternative  $\sigma$ -factor, as  $\sigma^{24}$  regulates expression of genes which

encode components of the extracytoplasmic (envelope) stress response (ESR).  $\sigma^{24}$ -mediated ESR becomes induced by various stressors, such as periplasmic misfolded proteins, non-optimal pH or oxidative stress and the ESR acts to adapt to and protect the bacterial envelope from these perturbations (Raivio, 2005). The majority of the work that has uncovered the modulation of  $\sigma^{24}$  expression and activity has been performed in E. coli but the number and organisation of orthologous genes within the rpoE operon are identical between E. coli and S. Typhimurium, thus the method of controlling RpoE expression in S. Typhimurium has been inferred (Rowley et al., 2006). The rpoE gene is encoded as part of an operon comprising the genes rseA, rseB and rseC. In S. Typhimurium rpoE is transcribed from 3 promoters, one of which is  $\sigma^{24}$ -dependent (Miticka et al., 2003).  $\sigma^{24}$  is under negative post-translational regulation during The inner membrane-spanning, RseA, is the cognate non-stress conditions. anti-sigma-factor for  $\sigma^{24}$ , acting by binding cytoplasmic RpoE to prevent  $\sigma^{24}$  interaction with RNAP, thus rendering  $\sigma^{24}$  inactive (Alba & Gross, 2004). Upon detection of extracytoplasmic stress, RseA is sequentially degraded by the inner membrane proteases DegS (cleavage at a periplasmic site) and YaeL/RseP (cleavage at a cytoplasmic site), releasing the cytoplasmic portion of RseA and  $\sigma^{24}$  into the cytoplasm. In the cytoplasm, the cleaved RseA and  $\sigma^{24}$  complex is directed towards ClpXP protease for final degradation of RseA to free  $\sigma^{24}$  to interact with RNAP and regulate expression of genes associated with the ESR (Rowley et al., 2006).

One mechanism by which members of the  $\sigma^{24}$  regulon contribute to the ESR is through sRNA-mediated degradation of OMP mRNAs upon induction of the ESR. Two sRNAs, RybB and MicA, perform this role in *S.* Typhimurium, while a third  $\sigma^{24}$ -dependent sRNA, MicL/RyeF, performs a similar role in *E. coli*. During growth under non-stress conditions the sRNAs are part of an auto-regulatory loop which limits OMP biogenesis to avoid accumulation of misfolded OMP intermediates in the periplasmic space (Papenfort *et al.*, 2006, Guo *et al.*, 2014). Prediction of RpoE binding sites across the genomes of ten Gram negative enteric bacterial species identified a core group of  $\sigma^{24}$ -dependent genes, involved in assembly and maintenance of LPS and OMPs, while *S.* Typhimurium-specific  $\sigma^{24}$ -dependent genes encode proteins important for a pathogenic lifestyle (Rhodius *et al.*, 2006). RpoE contributes to *S.* Typhimurium pathogenesis through regulating the transcription of genes required for survival of infection-associated

stresses, such as oxidative stress and resistance to antimicrobial peptides (Rowley et al., 2006). A  $\Delta rpoE$  mutant is attenuated for macrophage survival/replication compared to wild-type, presumably due to a reduction in expression of proteins involved in periplasmic protein folding and stability.  $\sigma^{24}$  is also required for full virulence in mice during systemic infection (Humphreys *et al.*, 1999). However the absence of the  $\sigma^{24}$  protein has no effect on the ability of *S.* Typhimurium to invade epithelial cells, indicating that much of the role of  $\sigma^{24}$  during *S.* Typhimurium pathogenesis can be attributed to survival within the intracellular compartments (Humphreys et al., 1999).

In this study, the  $\sigma^{24}$  regulon was investigated at LSP as this growth phase presents stresses known to induce the activity of  $\sigma^{24}$  and , thus, expression of genes within the  $\sigma^{24}$  regulon, which encode components the ESR (Rowley et al., 2006).

#### 1.2.10 PhoB/R

PhoB and PhoR comprise a TCS that responds to environmental phosphate levels and regulates target genes accordingly, to maintain phosphate homeostasis. In bacterial cells phosphate contributes to a variety of cellular functions, as a component of nucleotides which form DNA and RNA, as an energy store in the form of ATP, as an integral part of membranes in the form of phospholipids and in post-translational protein modifications, for example phosphorylation acts as a signal to activate proteins in TCSs (Chekabab *et al.*, 2014). Under phosphate limitation the sensor kinase, PhoR phosphorylates the response regulator PhoB, which activates the Pho regulon, including the genes involved in phosphate assimilation and transport. PhoB acts as a typical transcription initiator protein, by recruitment of RNAP via direct interaction with the σ-subunit (Chekabab et al., 2014). When phosphate is abundant the Pho regulon is not induced. The switch to the "off" state is mediated by an interaction between PhoR and the phosphate transporter PstSCAB-PhoU operon. PhoR acts as a phosphatase, maintaining PhoB in an unphosphorylated, and thus inactive, state (Chekabab et al., 2014).

Phosphate levels are limiting within the SCV, as reflected by the up-regulation of the *phoB/R* locus and the *pst* operon during macrophage infection (Eriksson *et al.*, 2003), as well as the activation of SPI2 genes by phosphate starvation (Deiwick *et al.*, 1999). It

has, therefore, been hypothesised that the PhoB/R TCS plays an important role in acquisition and transport of any available phosphate to aid in *S*. Typhimurium intracellular survival. PhoB plays an important role in virulence through repression of *hilA* transcription when phosphate levels are low. This is likely to be a mechanism of switching off invasion-associated genes, when they are no longer required, during intracellular infection (Lucas *et al.*, 2000). The mechanism of PhoB-mediated repression of *hilA* has been suggested to occur through activation of the negative regulators of SPI1, *fimYZ* and *hilE*, by PhoB (Jones, 2005).

In this study the PhoB/R regulon was investigated under SPI2-inducing conditions (growth to  $OD_{600}0.3$  in PCN medium with 0.4 mM phosphate, pH5.8 (Lober et al., 2006)) in order to focus on the role of the PhoB/R TCS under conditions which mimic the intracellular environment, when the PhoB/R regulon is expected to be active.

### 1.2.11 SlyA

SlyA belongs to the MarR family of transcriptional regulators, members of which are required for host adaptation in other bacterial pathogens (Stapleton *et al.*, 2002). SlyA was first identified as a haemolysin following *S*. Typhimurium growth on blood agar plates. SlyA is required for macrophage survival and for virulence in mice following i.p., intragastric (i.g.) and oral infection, however SlyA is not required for invasion of epithelial cells, but is required for the cytotoxicity of M cells and survival within Peyer's patches (Libby *et al.*, 1994, Daniels *et al.*, 1996). The *slyA* gene is highly expressed during macrophage infection, and absence of the SlyA protein renders *S*. Typhimurium hyper-susceptible to oxidative stress and the antimicrobial peptides polymyxin B and magainin 2, leading to the conclusion that SlyA is required for intracellular survival (Buchmeier *et al.*, 1997, Shi *et al.*, 2004). Expression of a number of SPI2 genes, including the genes encoding the primary SPI2 regulatory TCS SsrA/B, is dependent on SlyA (Linehan *et al.*, 2005, Navarre *et al.*, 2005) via SlyA binding to and directly regulating expression from the *ssrA* promoter (Okada *et al.*, 2007, Navarre et al., 2005).

Transcription of *slyA* is negatively auto-regulated and is activated by PhoP, the response regulator of the PhoP/Q TCS (Norte *et al.*, 2003, Stapleton et al., 2002). SlyA also mediates activation of *phoP* expression, forming a positive feedback loop (Song *et al.*,

2008). In addition to this regulatory loop, there is an overlap between genes that require both SlyA and PhoP for expression (Navarre et al., 2005). However, *S.* Typhimurium SlyA does not appear to act by direct transcription initiation via recruitment of, or interaction with, RNAP subunits. As previously discussed, SlyA acts to counter-silence repressive effects of H-NS and other NAPs (Stoebel et al., 2008). In addition to regulation of SPI2 genes and genes required for the intracellular survival of *S.* Typhimurium, two separate proteomic studies have identified the flagellin protein, FliC, as being positively regulated by SlyA, but the mechanism of this regulation has not been elucidated (Stapleton et al., 2002, Spory *et al.*, 2002).

Thus, SlyA is an important SPI2-associated TF. In this study the regulon of SlyA was investigated under SPI2-inducing conditions as SlyA and SlyA-target genes appear to be active under these conditions.

#### 1.2.12 SsrA/B

SsrA and SsrB comprise a SPI2-encoded regulatory TCS that is essential for expression of genes encoding components of the SPI2 TTSS apparatus and effectors, encoded on SPI2 and elsewhere on the chromosome (Cirillo et al., 1998, Tomljenovic-Berube et al., 2010). SsrA is the membrane-located sensor kinase, which is phosphorylated in response to an unknown environmental signal and subsequently phosphorylates and activates SsrB, the cognate response regulator. SsrB binds to all SPI2 gene promoters and is necessary for SPI2 expression and systemic Salmonella infection (Fass & Groisman, 2009). ssrA and ssrB are transcribed from separate promoters upstream of each gene, however differential RNA-seq (dRNA-seq) analysis, which determines the base position of genomic TSS (Sharma et al., 2010), has demonstrated that the ssrB promoter is approximately 3.5-fold more active than the ssrA promoter within murine macrophages (Srikumar et al., 2014). SsrB binds and regulates expression from its own promoter, as well as the ssrA promoter. H-NS silences expression from the ssrA promoter and, as previously mentioned, HilD mediates counter-silencing of this effect during stationary phase growth in rich media but not under conditions which mimic the intracellular environment (Bustamante et al., 2008). Under SPI2-inducing conditions SsrB has a dual function in antagonising the H-NS-mediated silencing of ssrA expression, as well as

classical transcription initiation (Walthers et al., 2007). In addition to H-NS, other NAPs, YdgT, Hha and Fis, directly and indirectly affect SPI2 gene expression through the SsrA/B TCS (Fass & Groisman, 2009). Enzyme II of the nitrogen-metabolic phosphotransferase system (EII<sup>Ntr</sup>) represses SsrB expression. However, this interaction is post-transcriptional and thought to occur through direct interaction between EII<sup>Ntr</sup> and the SsrB protein in order to avoid over-expression of SPI2 genes under SPI2-inducing conditions (Choi *et al.*, 2010). OmpR, the response regulator of the OmpR/EnvZ TCS binds and regulates expression from both the *ssrA* and *ssrB* promoters (Feng *et al.*, 2003). PhoP activates SPI2 gene expression, however in contrast to the direct activation of *ssrA* and *ssrB* transcription by OmpR, PhoP only binds and activates transcription from the *ssrB* promoter. PhoP activates *ssrA* expression post-transcriptionally, in a mechanism that requires the 5' UTR of the *ssrA* mRNA (Bijlsma & Groisman, 2005). As previously mentioned, SlyA directly affects SPI2 gene expression through direct activation of *ssrA* transcription (Okada et al., 2007).

The multiple regulatory inputs, both ancestral and horizontally-acquired, which are integrated at the *ssrA* and *ssrB* promoters permit the detection of appropriate environmental signals for up-regulation and co-ordinated expression of the genes required for intracellular survival and systemic infection. The SCV rapidly acidifies upon infection (Rathman *et al.*, 1996) and early induction of SPI2 genes under acidic conditions (pH 6.2 or lower) has been observed (Lober et al., 2006). An independent delayed induction, corresponding to entry into stationary phase, of SPI2 genes under phosphate starvation (0.4 mM) has also been demonstrated. Both of these environmental signals are integrated through the SsrA/B system (Lober et al., 2006). Acidic conditions increase OmpR binding to SPI promoters, which accounts, at least partially, for the increase in SPI2 gene expression at a low pH (Quinn *et al.*, 2014). Limiting concentrations of divalent cations, such as Mg<sup>2+</sup>, result in an increase in SPI2 gene expression, likely through sensing and regulation by the PhoP/Q TCS (Deiwick et al., 1999). Low levels of Ca<sup>2+</sup> and low osmolarity also induce SPI2 gene expression via activation of SsrA/B (Garmendia *et al.*, 2003).

In this study, the effect on gene expression of the absence of the SsrA protein and SsrB protein individually, as well as the absence of both proteins simultaneously was investigated under SPI2-inducing conditions.

## 1.2.13 PhoP/Q

The PhoP and PhoQ proteins are the respective response regulator and membrane-bound sensor kinase of the ancestral PhoP/Q TCS that is essential for S. Typhimurium virulence (Kato & Groisman, 2008). PhoP is required for S. Typhimurium virulence in mice and for intracellular survival. The absence of the PhoP or PhoQ proteins renders S. Typhimurium hypersensitive to antimicrobial peptides, acidic pH, Mg<sup>2+</sup> starvation and hydrogen peroxide (Kato & Groisman, 2008). PhoQ senses the concentration of extracellular divalent cations, such as Mg<sup>2+</sup> and Ca<sup>2+</sup>, and auto-phosphorylates under conditions of Mg<sup>2+</sup> and Ca<sup>2+</sup> starvation, subsequently phosphorylating and activating PhoP. It has also been suggested that antimicrobial peptides and acidic pH act as signals to activate the PhoP/O system (Bader et al., 2005, Prost et al., 2007), however there is some dispute about the mechanisms involved in PhoQ sensing of these signals (Groisman & Mouslim, 2006). Activated PhoP binds to the promoters of target genes, involved in Mg<sup>2+</sup> transport and virulence (Garcia Vescovi et al., 1996). PhoP contributes to S. Typhimurium virulence through differential regulation of both the SPI1 and SPI2 TTSSs (Kato & Groisman, 2008). PhoP represses the SPI1-encoded prg genes (PhoP Repressed Genes) likely through repression of hilA expression (Behlau & Miller, 1993, Bajaj et al., 1996). Repression of hilA expression by pag genes (PhoP activated Genes) requires an active HilD protein, suggesting that the negative effect of PhoP on SPI1 gene expression is mediated through HilD (Boddicker et al., 2003). PhoP also directly or indirectly represses transcription of the fliC gene, which encodes phase 1 flagellin (Adams et al., 2001).

In contrast to the PhoP-mediated repression of invasion-associated genes, which are no longer necessary under the intracellular conditions that activate the PhoP/Q TCS, PhoP activates expression of SPI2 genes and genes required for survival during the intracellular lifestyle of *S.* Typhimurium (Kato & Groisman, 2008). As previously mentioned, PhoP directly activates transcription from the *ssrB* promoter and activates *ssrA* via a post-transcriptional method (Bijlsma & Groisman, 2005). PhoP directly activates expression of the *sseL* gene, which encodes a SPI2-translocated effector protein (Gal-Mor *et al.*, 2011). However in the absence of the PhoP protein, SPI2 genes are activated to the same maximal expression level as in wild-type but the induction is highly delayed (Xu & Hensel, 2010). In order to mediate survival in the low Mg<sup>2+</sup> intracellular environment

PhoP activates expression of genes encoding the Mg2+ uptake systems, mgtA and mgtBCR, as well as other genes required for growth in low Mg<sup>2+</sup> (Soncini et al., 1996, Alix & Blanc-Potard, 2008). MgtC plays a role in intra-macrophage survival that is independent of the MgtC-mediated survival in low Mg<sup>2+</sup> conditions, indicating that MgtC has dual virulence-associated functions (Rang et al., 2007). PhoP activates expression of a long sRNA, AmgR, encoded antisense to the intergenic region between mgtB and mgtC, which is responsible for modulating the levels of the MgtB and MgtC proteins, and is, therefore, necessary for growth under low Mg<sup>2+</sup> conditions and virulence in mice (Lee & Groisman, 2010). Many PhoP-activated genes, such as pmrA/B, ugd, pagP and pagL, encode proteins which are required for re-modelling of the bacterial outer membrane through LPS modifications. These modifications contribute to the ability of S. Typhimurium to resist killing by host antimicrobial peptides and to evade the host immune system (Ernst et al., 2001). Additionally, PhoP contributes to  $\sigma^{38}$ -mediated transcriptional regulation by stabilisation of RpoS protein turnover during growth under low Mg<sup>2+</sup> conditions. This stabilisation is achieved through PhoP directly activating transcription of the iraP gene, the product of which interferes with presentation of the  $\sigma^{38}$ protein to the protease ClpXP (Tu et al., 2006).

As a result of the primary role of the PhoP/Q TCS in intracellular survival, PhoP and PhoQ are considered as SPI2-associated regulators. The PhoP/Q regulon was investigated under SPI2-inducing conditions in this study.

#### 1.2.14 *OmpR/EnvZ*

OmpR is the response regulator and EnvZ is the cognate membrane-bound sensor kinase of the ancestral OmpR/EnvZ TCS. In *E. coli*, high osmolarity results in an induction of OmpR expression, however in *S.* Typhimurium high osmolarity does not cause induction of OmpR. Instead, acid shock induces *S.* Typhimurium OmpR expression (Bang *et al.*, 2002, Quinn et al., 2014). The *ompR* promoter is subject to silencing by H-NS, but OmpR binds to its own promoter and activates *ompR* expression via a counter-silencing mechanism, likely through acid-induced changes in DNA supercoiling around the *ompR* promoter (Bang et al., 2002). Phosphorylated OmpR binds and activates transcription from the promoters of genes required for the acid tolerance response (ATR) to pre-adapt

the bacteria to more severe acid shocks (Foster & Hall, 1990). In addition to the role of OmpR in acid shock resistance, OmpR plays a role in S. Typhimurium virulence, as mutation of the *ompR* gene causes attenuation following oral and intravenous infection of mice (Dorman et al., 1989). As previously discussed, OmpR is necessary for activation of SPI2 genes via transcriptional control of the ssrA and ssrB promoters (Feng et al., 2003). OmpR appears to be essential for transcription initiation of ssrA even in the absence of H-NS, therefore OmpR does not act simply by counter-silencing H-NS but probably acts by recruiting RNAP to the ssrA promoter (Bustamante et al., 2008). OmpR binding to the ssrA promoter is increased under acidic conditions and by relaxation of DNA supercoiling (Cameron & Dorman, 2012, Quinn et al., 2014). Furthermore, OmpR and the NAP, Fis, have been implicated in transcriptional priming of SPI2 genes while in the host intestinal lumen. Transcriptional priming is SsrB-independent, instead OmpR binds the ssrA promoter under non-SPI2-inducing conditions and activates a low level of ssrA expression. This mechanism is thought to prepare the cells for entry into the intracellular vacuole as the SPI2 TTSS is required immediately upon internalisation (Osborne & Coombes, 2011).

OmpR activates transcription of SPI1 genes by a mechanism which is thought to involve the HilD protein, because *hilC* transcription is reduced in the absence of the OmpR protein while transcription of *hilD* is not affected in the absence of the OmpR protein (Lucas & Lee, 2001, Ellermeier et al., 2005). More recently it has been demonstrated, however, that OmpR binds to both the *hilC* and *hilD* promoters when DNA supercoiling is relaxed and under acidic conditions, inducing transcription of the former and repressing transcription of the latter (Cameron & Dorman, 2012, Quinn et al., 2014). OmpR binds the promoters of genes encoding SPI1- and SPI2-translocated effector proteins, encoded outside of SPI1 and SPI2. OmpR also directly regulates expression of genes that encode important virulence-associated structures, such as the SPI4-encoded non-fimbrial adhesin, flagella, fimbriae and curli fibres (Quinn et al., 2014, Perkins *et al.*, 2013). There is a strong overlap between the OmpR and PhoP regulons, as well as OmpR binding and regulating expression from the promoter of the *phoP* gene (Quinn et al., 2014). These regulatory interactions highlight the integral role played by OmpR in *S.* Typhimurium virulence, in conjunction with OmpR regulation of the key SPI1 and SPI2 regulators.

Thus OmpR can be viewed as a global regulator of *S*. Typhimurium virulence. In this study we have focussed on the OmpR/EnvZ regulon under SPI2-inducing conditions, which involves growth at pH 5.8, as the OmpR protein and OmpR-target genes are highly expressed and active under these conditions.

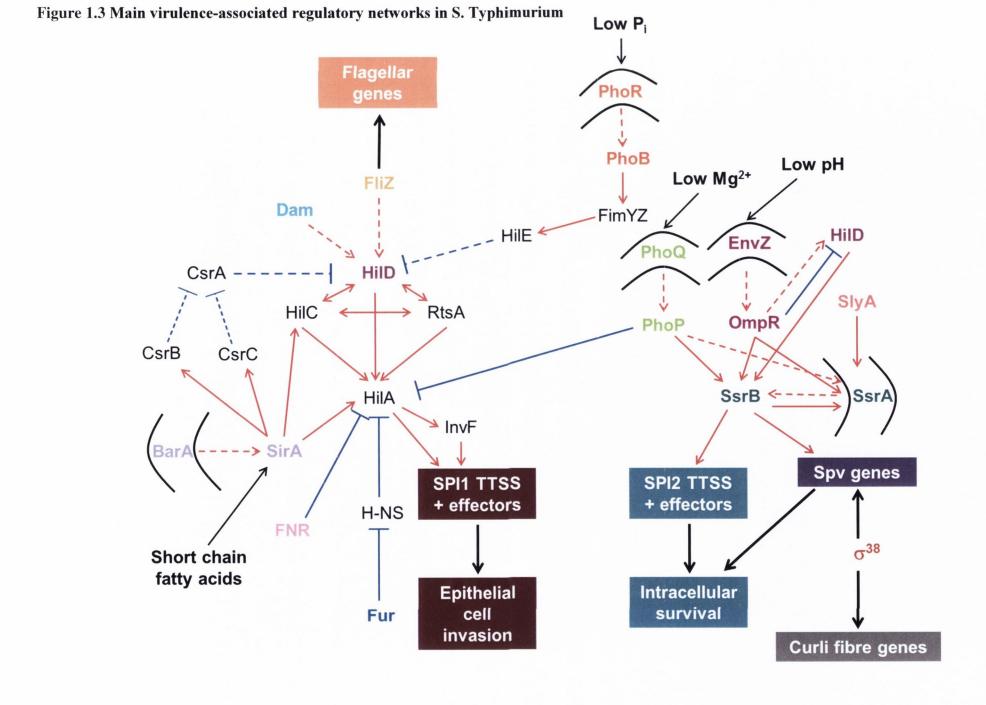
#### 1.2.15 FNR

Enteric bacteria, such as E. coli and S. Typhimurium, experience oxygen fluctuations at various sites within the host. For example, oxygen is in limited supply in the lumen of the small intestine but there is an oxygenated zone adjacent to the gastrointestinal mucosa (Marteyn et al., 2010). In facultative anaerobes, under oxygen limiting conditions, oxygen utilisation systems must be shut down, and cells produce energy by anaerobic respiration, using alternative electron acceptors such as fumarate and nitrate, or by fermentation of sugars (Salmon et al., 2003). The E. coli FNR (Fumarate and Nitrate Reductase) protein was first identified in a screen for mutants unable to reduce fumarate and nitrate (Lambden & Guest, 1976). S. Typhimurium FNR (originally named OxrA) (Jamieson & Higgins, 1986) is a global regulator that senses changes in oxygen concentration through an iron-sulphur cluster (4Fe-4S). FNR is active as a dimer containing the iron-sulphur cluster, under anaerobic conditions. In the presence of oxygen the iron-sulphur cluster dissociates, resulting in a de-stabilisation of the FNR dimer and a loss of FNR activity (Kiley & Beinert, 2003). Gene products required during the transition between aerobic and anaerobic growth, and for anaerobic respiration, are typically regulated at the transcriptional level by FNR (Fink et al., 2007).

The FNR regulon has been identified in both *E. coli* and in *S.* Typhimurium and there is an overlap between FNR functionality in the two species. The core FNR regulon from both species comprises genes involved core metabolic function, energy metabolism and motility but *S.* Typhimurium-specific FNR function involves the regulation of genes required for virulence (Fink et al., 2007). Expression of SPI1 genes is reduced in the absence of the FNR protein under anaerobic conditions. This interaction has not been demonstrated to be a direct effect but low oxygen induces expression of SPI1 genes and this signal could be integrated, at least partially, by FNR. However, the regulatory link between FNR and SPI1 genes is likely to be complex, as expression of a *hilA* 

transcriptional fusion is increased in the absence of FNR under SPI1-inducing conditions (Golubeva *et al.*, 2012). Interestingly, an *fnr* mutant is highly attenuated for virulence in mice following oral and i.p. infection, and is also highly attenuated for survival in murine macrophages (Fink et al., 2007). The *fnr* mutant is attenuated for virulence following infection of iNOS<sup>-/-</sup> mice, which are deficient in producing RNS but the *fnr* mutant was not attenuated in mice which are deficient in producing ROS, indicating that FNR controls genes involved in intracellular survival of ROS but not RNS (Fink et al., 2007).

As a result of the importance of FNR for *S*. Typhimurium survival under anaerobic conditions, the growth rate of the  $\Delta fnr$  mutant is greatly impaired compared to wild-type when grown under *in vitro* conditions known to activate fnr and genes of the FNR regulon. We, therefore, investigated the FNR regulon using a defined minimal medium, MMA (Table 2.1 and Table 2.3) (Miller, 1972), as preliminary data showed that the  $\Delta fnr$  mutant grew as well as the wild-type in this medium (Appendix VIII). Glycerol was used as a carbon source to avoid glucose-mediated catabolite repression at some FNR-dependent promoters. The use of the non-fermentable glycerol carbon source allows the growth of an fnr mutant in the presence of fumarate and trimethylamine N-oxide (TMAO) as alternative terminal electron acceptors (Constantinidou *et al.*, 2006). In this study the wild-type and  $\Delta fnr$  mutant were cultured in the fumarate and TMAO-supplemented MMA medium, without aeration (static growth) to an OD<sub>600</sub>0.3. The activity of the FNR regulon was assessed based on expression of the anaerobically-inducible FNR-dependent sRNA, FnrS (Boysen *et al.*, 2010, Durand & Storz, 2010), as determined by northern blot (Appendix VIII).



## Figure 1.3 Main virulence-associated regulatory networks in S. Typhimurium

Schematic showing the main virulence-associated regulatory pathways in *S.* Typhimurium. Some interactions have been omitted for ease of understanding. Regulators that are relevant for this study are labelled in colour. The membrane-bound sensor kinase of each TCS, is shown within 2 curved black lines, representing the membrane. Some of the main environmental stimuli (low phosphate, low magnesium, low pH, short chain fatty acids) sensed by TCS components are indicated. Full red arrows denote transcriptional activation, dashed red arrows denote post-transcriptional activation. Full blue T-bars denote transcriptional repression, dashed blue T-bars denote post-transcriptional repression. Thick black arrows are used to represent general regulation of groups of genes and cellular processes (indicated in coloured boxes), rather than specific interactions.

## 1.3 Small non-coding RNAs (sRNAs)

#### 1.3.1 Overview

Protein factors are not the only important players in regulation of gene expression. In addition to transcription initiation, it is now recognised that post-transcriptional regulation of gene expression by various RNA species, such as sRNAs and cis-encoded RNA elements, is an important control point in ensuring appropriate gene expression (Altuvia & Wagner, 2000). sRNAs are typically transcribed from intergenic regions (IGR) of the chromosome, that were originally thought to be empty or transcriptionally silent (Vogel, 2009). However various studies that employed computational predictions based on sequence conservation in IGRs between related bacteria, and the presence of transcription signals, such as orphan promoters upstream of termination sequences, led to discovery of many sRNAs (Altuvia, 2007). Other approaches including DNA microarrays containing probes for both strands of the IGR, shot-gun cloning and co-precipitation of sRNA molecules with RNA chaperone proteins have also been useful tools in sRNA discovery (Altuvia, 2007). In recent years there have been many technological advances and increased accessibility of high-throughput sequencing technologies, such as RNA-seq. RNA-seq involves the sequencing of cDNA, allowing transcriptome profiling with single nucleotide resolution (Ozsolak & Milos, 2011). High-throughput technologies, coupled with some of the previously mentioned sRNA discovery tools, or with the use of total RNA extracts, have led to the identification of rare or low-abundance sRNA molecules under a variety of environmental conditions. This has substantially increased our knowledge about the expression patterns and potential functions of previously identified

sRNAs, in addition to identifying novel sRNAs (Sittka *et al.*, 2008, Chao et al., 2012, Kröger et al., 2012, Kröger et al., 2013).

The E. coli and S. Typhimurium model systems have been at the forefront of investigations of sRNA function, and examples of sRNA-mediated gene regulation will be provided from both organisms. sRNAs can range in length from 50 to 500 nt (Altuvia, 2007), although some longer examples have been reported, for example the S. Typhimurium antisense-encoded AmgR sRNA is 1.2 kb in length (Lee & Groisman, sRNAs are usually non-coding and exhibit a regulatory function, typically post-transcriptionally. sRNAs can be encoded in cis or in trans to their target gene (Vogel, 2009). The group of cis-encoded sRNAs comprise riboswitches, RNA thermometers and antisense-encoded sRNAs. Riboswitches are encoded 5' to their target genes and regulate transcription or translation of the downstream gene in response to environmental signals or metabolites. RNA thermometers are inhibitory structures which sequester the ribosome binding site (RBS) of the downstream gene and prevent translation. RNA thermometers are highly temperature-sensitive and can thus mediate temperature-associated translational control. Antisense-encoded sRNAs are transcribed on the negative strand of a coding gene. The gene on the opposite strand is often regulated by the antisense sRNA, as a result of base-pairing with perfect sequence complementarity (Vogel, 2009).

Unlike antisense-encoded sRNAs, *trans*-encoded sRNAs act at a distance and typically function via the formation of a short imperfect base-pairing interaction with the mRNA of their target gene. To date, most of the reported functional *trans*-acting sRNAs require the RNA chaperone protein Hfq to maintain stability of the sRNA prior to target binding, to aid in target binding by the sRNA and to stabilise the interaction between the sRNA and target mRNA (Gottesman, 2005). The canonical model of sRNA-mediated gene regulation involves binding of the sRNA at the 5' end of the target message, typically between -20 nt and +15 nt relative to the first nucleotide of the start codon of the target gene (Papenfort *et al.*, 2010). sRNA-mediated base-pairing in this region occludes the ribosome binding site (RBS) and blocks binding by the 30S ribosomal subunit, thereby preventing translation initiation. This mechanism usually results in the de-stabilisation of the target message by the actions of cellular ribonucleases (Gottesman, 2005). However,

non-canonical mechanisms of sRNA-mediated gene regulation are regularly elucidated, highlighting the flexibility of these riboregulators (Desnoyers *et al.*, 2013) (Figure 1.4).

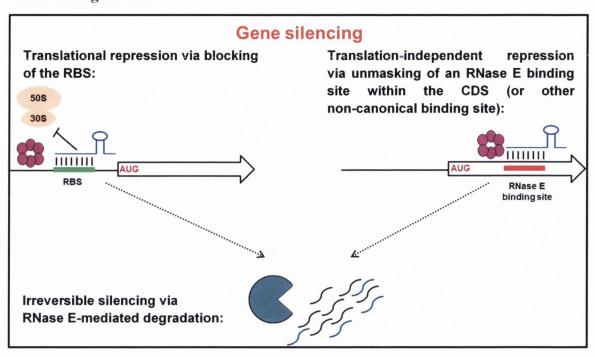
The need for non-canonical methods of sRNA-mediated gene regulation often arises from sequence or structural features in the target mRNA (Desnoyers et al., 2013). Some sRNAs have been shown to activate, rather than repress target gene expression. sRNAs that act as activators of gene expression will be discussed in more detail in section 5.1.2. Non-canonical methods of sRNA-mediated translational silencing have been reported, such as, binding of the sRNA at sites other than the 5' end of the transcript. The sRNA IstR-1 inhibits translation of its target mRNA, the toxin-encoding tisAB, through binding at a site far upstream of the translation initiation region (TIR) for tisAB. The site bound by IstR-1 is thought to act as a standby site for ribosomes. Ribosome standby sites are single stranded regions of DNA adjacent to the TIR. Standby sites are required upstream of genes which are not efficiently translated as a result of inhibitory structures at the RBS. Ribosomes bind to the standby site and may slide into the RBS when the inhibitory structures transiently unfold. IstR-1 binds to the standby site upstream of the tisAB mRNA, making this region double-stranded and preventing ribosome binding (Darfeuille et al., 2007). Binding of OmrA and OmrB far upstream of the csgD RBS, results in repression of csgD translation through an unknown mechanism (Holmqvist et al., 2010). A further example, of a non-canonical sRNA binding site leading to a non-canonical mechanism of sRNA-mediated translational control, involves the sRNA Spot42. The Spot42 binding site is too far upstream of the TIR of the sdhC mRNA, thus Spot42 does not directly interfere with ribosome binding. Instead Spot42 acts as a recruitment factor for the RNA chaperone, Hfq, which binds within the sdhC TIR and competes with the 30S ribosomal subunit (Desnoyers & Masse, 2012). A further example, of the non-canonical role of Spot42 in regulation of the sdhCDAB mRNA, is that Spot42 binding does not result in degradation of the sdhCADB mRNA. It has been suggested that Spot42 has limited regulatory impact upon sdhCADB translation, and that residual ribosomes protect the sdhCADB mRNA from degradation (Desnoyers & Masse, 2012). Furthermore, sRNAs, such as GcvB, bind to translation enhancer sequences in target mRNA resulting in a shut-down of translation (Sharma et al., 2007).

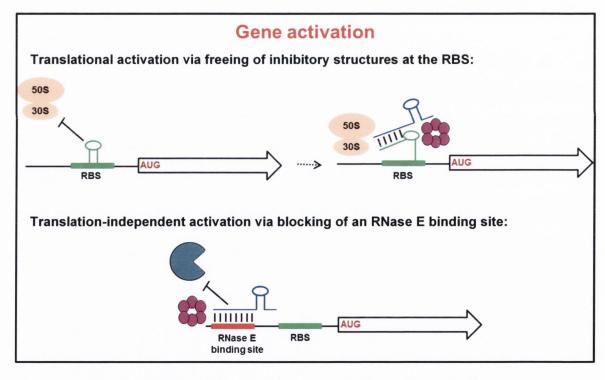
In contrast, translation-independent mechanisms of sRNA-mediated post-transcriptional gene regulation have also been reported. The activation of a long isoform of the fatty

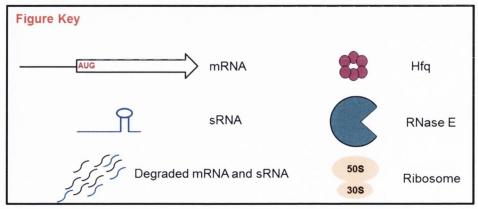
acid synthase-encoding gene, cfa, by the sRNA, RydC, involves RydC binding to a non-canonical site upstream of the cfa RBS and protecting cfa from degradation by the endoribonuclease RNase E, without affecting cfa translation (Frohlich et al., 2013). Furthermore, two sRNAs, MicC and MicF bind their respective mRNA targets, ompD and lpxR, within the coding sequence (CDS) of the transcript. It had been considered that sRNAs would not be able to repress target translation through binding deep within the target CDS, because the helicase activity of the 70S ribosome could unwind sRNA-target Instead, MicC and MicF appear to act by accelerating base-pairing interactions. ribonuclease-mediated target decay, perhaps through unmasking an RNase E binding site (Pfeiffer et al., 2009, Corcoran et al., 2012). Similar mechanisms have been suggested for SgrS activation of yigL (Papenfort et al., 2013) and SdsR repression of ompD (Frohlich et al., 2012). Furthermore, in a screen for RybB binding sites within target mRNAs, it was shown that RybB targets sites within an approximately 100 bp window on either side of the first nucleotide of the initiation codon, suggesting that RybB may act via translation-dependent or translation-independent mechanisms, as well as nucleolytic or non-nucleolytic mechanisms (Papenfort et al., 2010). A final example of non-canonical sRNA-target binding involves the sRNA DsrA, which binds a site at the stop codon of the hns transcript, as well as a canonical binding site near the hns start codon. Simultaneous binding by DsrA to both sites results in the formation of a contiguous coaxial stack, which loops out the internal part of the hns mRNA, making this region more accessible to ribonuclease-mediated degradation (Lease & Belfort, 2000).

Finally sRNAs can function to directly affect protein activity. As previously discussed, the sRNAs CsrB and CsrC redundantly bind and titrate the RNA-binding CsrA protein from its targets (Fortune et al., 2006). In *E. coli*, the sRNA McaS also sequesters CsrA, preventing CsrA binding to and repressing expression of *pgaA*, which is necessary for biofilm formation. McaS presents an interesting example as, in addition to directly targeting the CsrA protein, McaS can act as a base-pairing sRNA (Jorgensen *et al.*, 2013).

Figure 1.4 Main mechanisms of post-transcriptional regulation by Hfq-bound *trans*-acting sRNAs







# Figure 1.4 Main mechanisms of post-transcriptional regulation by Hfq-bound trans-acting sRNAs

Schematic showing the most common mechanisms of action of Hfq-bound *trans*-acting base-pairing sRNAs. The canonical method of sRNA-mediated regulation of gene expression involves binding of the sRNA within the translation initiation region (TIR) of the target mRNA, preventing ribosome binding. This process is usually accompanied by ribonuclease-mediated irreversible gene silencing i.e. degradation. Non-canonical sRNA binding sites, upstream of the TIR or within the coding sequence of the gene, are usually regulated via translation-independent mechanisms, which typically unmask ribonuclease binding sites and accelerate degradation. sRNA-mediated translation activation may occur at genes which contain inhibitory structures within the TIR. Base-pairing by the sRNA may relieve the inhibitory structure, allowing ribosome binding. Translation-independent mechanisms of sRNA-mediated activation may also occur. Binding of sRNAs to ribonuclease binding sites outside of the TIR may block ribonuclease binding.

## 1.3.2 Cellular functions of trans-acting sRNAs

The benefits to the cell of using sRNA-based methods of gene regulation are becoming appreciated with every new mechanism that is elucidated. It has been suggested that every stress response will contain at least one sRNA within the regulon (Gottesman, 2005). The advantages of sRNA-mediated regulation of stress responses are obvious, as de novo synthesis of sRNAs is rapid in comparison to de novo protein synthesis. Therefore, sRNAs can quickly respond to the stress signal and up- or down-regulate genes accordingly (Beisel & Storz, 2010). Not only are sRNAs themselves synthesised more rapidly but sRNA-mediated regulation is likely to result in the desired regulatory effect more rapidly. Post-transcriptional regulation is a step closer to protein production than transcriptional regulation, thus sRNA-mediated targeting of mRNAs results in a more immediate effect on protein output (Shimoni et al., 2007). Stress-induced sRNAs are also likely to be highly stable under the conditions when they are induced and can be rapidly degraded when the sRNA function is no longer required (Altuvia & Wagner, 2000, Beisel & Storz, 2010). The production of sRNAs is also likely to be more economical for the cell. The small size of sRNAs and their lack of requirement for translation machinery indicate that production of sRNAs is more energetically cost-effective than the production of proteins, and energy conservation is likely to be a key consideration for cells under stress (Beisel & Storz, 2010). Finally the previously mentioned versatility and adaptability of sRNA molecules to mediate regulatory function, using a diverse range of mechanisms and targets, is likely to be one of the most important advantages of RNA-mediated regulation (Desnoyers et al., 2013). TFs typically initiate or repress transcription through interaction with RNAP or through re-modelling of promoters to affect RNAP promoter recognition and binding (Lee et al., 2012). However, sRNAs appear to have adopted a number of mechanisms of action, from interacting with ribosomal subunits, base-pairing with structured RNA to alleviate or enhance *cis*-acting riboregulation, interacting with proteins and interacting with the cellular degradation machinery (Desnoyers & Masse, 2012).

These benefits of sRNA-mediated gene regulation, over protein-mediated gene regulation, have resulted in the integration of sRNAs into cellular regulatory networks, resulting in "mixed regulatory networks". A number of sRNAs exert a global regulatory effect on the cell through targeting TF mRNAs, creating highly interconnected cellular regulatory networks, in which transcriptional and post-transcriptional regulatory nodes act together to mediate cellular genetic output (Beisel & Storz, 2010, Gopel & Gorke, 2012, Mandin & Guillier, 2013). Mixed regulatory systems are likely to provide tighter regulatory control of target genes and may integrate additional regulatory signals. There have been a number of reported cases of mixed regulatory loops, which can maintain cellular homeostasis or amplify regulatory signals (Beisel & Storz, 2010). The discovery that the sRNA, MicF, represses expression of the global TF, Lrp, has led to the suggestion that the majority of genes in a bacterium are under the direct or indirect control of an sRNA (Holmqvist *et al.*, 2012, Corcoran et al., 2012). Further examples of sRNAs that target TF mRNAs will be discussed below.

The role of sRNAs in regulation of stress responses has been well-documented (Vogel, 2009). The large number of sRNAs involved in regulation of OMP biogenesis highlights the importance of sRNAs in maintaining outer membrane homeostasis, particularly under stressful conditions (Mizuno *et al.*, 1984, Papenfort et al., 2006, Guillier & Gottesman, 2006, Pfeiffer et al., 2007, Papenfort *et al.*, 2008, Pfeiffer et al., 2009, Frohlich et al., 2012). In addition to environmental conditions which induce membrane stress, there are other examples of stress-induced sRNAs that play a role in bacterial survival and homeostasis of cellular systems. Resistance to oxidative stress and hydrogen peroxide-induced DNA damage is mediated by the sRNA, OxyS (Altuvia et al., 1997).

The PhoP/Q-dependent sRNA MgrR represses expression of *eptB* which encodes a protein involved in LPS modification. MgrR contributes to the *E. coli* PhoP/Q regulon by playing a role in resistance to antimicrobial peptides, such as polymyxin B, as cell sensitivity to antimicrobial peptides is mediated by LPS (Moon & Gottesman, 2009). Anaerobic shock induces expression of the sRNA FnrS, which regulates genes involved in anaerobic metabolism and survival during anaerobic conditions (Durand & Storz, 2010, Boysen et al., 2010). RyhB maintains iron homeostasis to prevent iron scarcity or toxicity within the cell (Masse et al., 2005) and SgrS regulates glucose homeostasis to maintain energy for metabolic processes while avoiding phosphosugar toxicity (Vanderpool & Gottesman, 2004, Papenfort et al., 2013). The sRNA GcvB contributes to cell survival through global regulation of genes involved in amino acid uptake and metabolism (Sharma *et al.*, 2011). The positive regulators of  $\sigma^{38}$  expression, DsrA (induced under cold-shock), RprA (induced by the RcsC/RcsB phosphorelay system) and ArcZ (induced under aerobic conditions) are integral parts of the *E. coli* stress response (Majdalani et al., 1998, Majdalani et al., 2002, Mandin & Gottesman, 2010).

Many of the sRNAs involved in mediating stress responses and bacterial survival play important roles during pathogenesis because many of the inducing stresses, such as oxidative stress, anaerobiosis, exposure to antimicrobial peptides, iron limitation etc. are encountered within the host, during infection. However, some sRNAs play a direct role in virulence through targeting specific virulence factors (Romby *et al.*, 2006, Hebrard *et al.*, 2012). One of the first hints that sRNAs were involved in *S.* Typhimurium virulence was that a mutant lacking the RNA chaperone Hfq is greatly attenuated for invasion of epithelial cells, intracellular survival and virulence in mice, and Hfq is required for expression and secretion of important virulence factors. This suggested that some of the Hfq-dependent sRNAs are involved in these virulence functions (Sittka *et al.*, 2007).

Three *S.* Typhimurium sRNA mutants were identified in a screen for genes that affect fitness during competitive infection in mice. Absence of IstR and SroA led to a fitness decrease, while absence of OxyS led to a fitness increase, compared to wild-type (Santiviago *et al.*, 2009). The functions of these sRNAs have not been elucidated in *S.* Typhimurium but the *E. coli* orthologues of each sRNA have been characterised. *E. coli* IstR-1 is part of a toxin/antitoxin system with the toxin, TisA/B. Expression of *tisA/B* is induced in response to DNA damage and TisA/B causes cellular toxicity. IstR-1 is

constitutively expressed, presumably to restrict TisA/B-induced toxicity during normal growth (Vogel *et al.*, 2004, Darfeuille et al., 2007). SroA is thought to act as a riboswitch for the *thiBPQ* thiamine transport operon in *E. coli* (Vogel *et al.*, 2003) and is also up-regulated in *S.* Typhimurium cells challenged with tigecylcine and tetracycline (Yu & Schneiders, 2012). As previously mentioned, *E. coli* OxyS is involved in resistance to oxidative damage. OxyS is, in fact, a pleiotropic regulator that may regulate as many as 40 genes in *E. coli*, including repression of the genes encoding the transcriptional regulators FlhA and  $\sigma^{38}$  (Altuvia et al., 1997).

A predictive algorithm was used to identify sRNAs encoded with *S*. Typhimurium specific islands, to search for sRNAs which may play a role in *S*. Typhimurium virulence, as opposed to sRNAs which are conserved between *E. coli* and *Salmonella* (Padalon-Brauch *et al.*, 2008). Nineteen island-encoded sRNAs were identified. Expression of many of these sRNAs increases during macrophage infection and two of the sRNAs have been shown to play direct roles in *S*. Typhimurium virulence. The expression of the island-encoded *isrJ* gene is lower in the absence of the HilA protein, and a strain deficient in IsrJ demonstrates reduced invasion rates of epithelial cells and reduced translocation of effector proteins into HeLa cells (Padalon-Brauch et al., 2008). The island-encoded IsrM sRNA is important for *S*. Typhimurium invasion of epithelial cells, intra-macrophage survival and replication and virulence in mice. IsrM independently represses expression of the SPI1 negative regulator HilE and the SPI1 effector protein SopA, through binding around the Shine-Dalgarno sequences of each mRNA. The loss of HilE and SopA expression in a Δ*isrM* mutant results in dysregulation of a number of SPI1 genes, accounting for the virulence defect of the mutant strain (Gong *et al.*, 2011).

In S. Typhimurium, SgrS represses translation of the SPI1- and SPI2-translocated effector protein SopD. SgrS does not regulate the almost identical SopD2 effector protein, due to a single nt difference between sopD and sopD2 in the putative SgrS interaction site. This represents an interesting example of co-option of an ancestral sRNA gene function to regulate a horizontally-acquired gene. As SgrS typically acts to modulate sugar homeostasis, it has been suggested that SgrS-mediated regulation of SopD expression may be a mechanism to detect host cell-type, based on carbon availability (Papenfort et al., 2012).

The regulator of the genes encoding curli fibres, CsgD, has proved to be a hub for sRNA-mediated gene regulation (Jorgensen *et al.*, 2012). Curli fibres are necessary for biofilm formation. Biofilms are a key virulence trait of *S*. Typhimurium as the multicellular lifestyle within a biofilm promotes bacterial survival under harsh environmental conditions or nutrient limitation (Fabrega & Vila, 2013). SdsR and ArcZ play a role in the multicellular lifestyle of *S*. Typhimurium through positive regulation of the *csgD* mRNA (Monteiro *et al.*, 2012). In contrast, a number of sRNAs down-regulate curli production through interaction with the *csgD* 5' UTR in *E. coli.* sRNAs in this latter group include OmrA and OmrB (Holmqvist et al., 2010), McaS (Thomason *et al.*, 2012, Jorgensen et al., 2012), GcvB and RprA (Jorgensen et al., 2012).

The important role of CsrB and CsrC in regulating invasion gene expression has already been discussed (section 1.2.5) (Fortune et al., 2006). However, both sRNAs have also been implicated in control of fimbrial gene regulation in *S.* Typhimurium. Fimbrial gene expression is under tight hierarchical control. The plasmid-encoded fimbrial operon *pef* is only expressed after expression of the chromocomal *fim* operon has ceased (Sterzenbach *et al.*, 2013). Expression of *pef* requires CsrA binding to activate expression. CsrB and CsrC sRNAs co-operate with the 5' UTR of the *fim* operon to antagonise CsrA function, preventing expression of the *pef* genes at the incorrect time (Sterzenbach et al., 2013).

These examples indicate the important role played by sRNAs in survival of harsh conditions experienced during the infection process, as well as in regulation of expression of genes and factors directly required for virulence. Many *Salmonella*-specific sRNAs have been identified recently and their characterisation should elucidate many more virulence-associated sRNAs (Kröger et al., 2012, Kröger et al., 2013).

### 1.3.3 Hfq

Hfq was originally identified in *E. coli*, as an essential host factor of the RNA bacteriophage Qβ. Hfq is an RNA-binding protein that is a core component of bacterial post-transcriptional regulatory networks and co-immunoprecipitation with Hfq has proved to be a useful tool in the discovery of new *trans*-acting sRNAs in *E. coli* and *S.* Typhimurium (Zhang *et al.*, 2003, Sittka et al., 2008, Chao et al., 2012). The Hfq protein

has a hexameric, ring-like structure that facilitates the binding of *trans*-acting sRNAs with their target mRNAs through short imperfect base-pairing interactions (Vogel & Luisi, 2011). The Hfq ring-like structure provides 2 RNA binding faces, proximal and distal, each with a specific sequence preference to allow Hfq to bring 2 different RNA molecules together. As a result of the hexameric nature Hfq, each face is repetitive in sequence to allow co-operative binding of many transcripts of the same RNA, as well as competitive binding by other RNAs (Vogel & Luisi, 2011). Recently a third surface of the Hfq hexamer for sRNA binding, the lateral surface, has been described. The current model for Hfq binding of sRNAs suggests that Hfq recognises and binds the rho ( $\rho$ )-independent terminator of sRNAs. The 3' end of the sRNA is then anchored on the proximal face of Hfq, while internal nucleotides of the sRNA interact with the lateral face of Hfq, stabilising the complex (Sauer, 2013).

Hfq stabilises bound sRNAs, protecting the sRNAs from degradation by cellular RNases, prior to target binding and the exertion of regulatory function (Moll et al., 2003). In most documented cases it appears that Hfq mainly acts as a catalyst to bring the sRNA and mRNA into contact with each other, and Hfq then dissociates from the complex, not playing a direct role in the regulatory function (Vogel & Luisi, 2011). The numbers of molecules of Hfq within the cell are limiting in comparison to the abundance of RNA species that bind Hfq. However, the pool of available Hfq molecules does not become sequestered or exhausted. A dynamic RNA-cycling model has been suggested as the method by which Hfq can simultaneously bind multiple abundant RNA species. The cycling model posits that RNA molecules are continuously binding and dissociating from Hfq as a result of transient binding by competitor RNAs (Fender et al., 2010). Hfq can associate with other proteins and protein complexes to play a role in control of gene expression. Hfq associates with the main catalytic component of the RNA degradosome, endoribonuclease RNase E, possibly as a means of coupling sRNA-mediated translational repression and irreversible target degradation. In addition, Hfq co-purifies with other proteins associated with RNA degradation, such as polynucleotide phosphorylase (PNPase) and poly(A) polymerase (PAP I) (Vogel & Luisi, 2011).

Hfq affects gene regulation in a number of ways, depending on the function of the individual sRNA to which Hfq is bound. As previously mentioned, an S. Typhimurium mutant strain lacking Hfq is severely attenuated for invasion, intracellular replication,

virulence in mice and protein secretion. The  $\Delta hfq$  mutant is also non-motile, due to reduced flagellin production, and is undergoing chronic envelope stress, as a result of the accumulation of mis-folded periplasmic proteins. Approximately 1/5 of all S. Typhimurium genes are differentially expressed in the absence of the Hfq protein, highlighting the importance of Hfq in indirect and direct regulation of gene expression (Sittka et al., 2007, Sittka et al., 2008). The Hfq regulon was investigated under SPI1-inducing conditions (ESP) in this study.

#### 1.3.4 PNPase

Endonucleolytic and exonucleolytic RNases are responsible for the degradation of mRNA and sRNA transcripts and are, therefore, important post-transcriptional regulators of gene expression. The RNA degradosome is a multi-protein complex involved in the degradation of RNA. The main components of the RNA degradosome are RNase E (an endonuclease), PNPase (an exonuclease), RhlB (a helicase) and an enolase of unknown function within the degradosome. The combined functions and co-operation of multiple enzymes within the degradosome allow for the degradation of highly structured RNA molecules (Carpousis, 2007). Other cellular ribonucleases include RNase G, RNase III, RNase P, RNase Z, RNase II and RNase R. Each ribonuclease has a different specificity and preference for target RNA species. The levels of RNases can be modulated by each other or via environmental signals to ensure that the genes which they control are only degraded under appropriate conditions (Viegas & Arraiano, 2008).

PNPase is a 3' exoribonuclease, which acts by using phosphate to cleave phosphodiester bonds (phosphorolysis), resulting in the processive release of nucleotides from RNA in a 3' to 5' direction (Carpousis, 2007). PNPase is encoded from the *pnp* gene, expression of which is negatively auto-regulated at the post-transcriptional level (Carzaniga *et al.*, 2009). In *E. coli* PNPase is essential for growth under low-temperature conditions and PNPase also contributes to cold acclimatisation in *S.* Typhimurium (Rouf *et al.*, 2011). A 3' overhang of unpaired nucleotides is required for an RNA molecule to be bound by PNPase. Poly(A) polymerase I, therefore, aids in exonucleolytic cleavage of RNA transcripts by polyadenylating the stable 3' end of transcripts, providing a "toe-hold" for

exonucleases, such as PNPase, that cannot readily digest complex secondary structures (Carpousis, 2007).

PNPase contributes to *S.* Typhimurium virulence. A single point mutation in *S.* Typhimurium *pnp* leads to an increase in SPI1 and SPI2 gene expression and a concomitant increase in invasion of epithelial cells and survival within murine macrophages, as well as the establishment of persistent infection in BALB/c mice (Clements *et al.*, 2002). Growth of a PNPase deficient strain under SPI2-inducing conditions and within murine macrophages demonstrated the link between PNPase and the plasmid-encoded *spv* virulence operon. *spv* gene expression is increased in the absence of PNPase in a SpvR-dependent fashion. The increased competitive fitness of a strain deficient in PNPase, compared to wild-type, is also dependent on SpvR (Ygberg *et al.*, 2006).

In this study the role of PNPase was investigated under SPI1-inducing conditions to further investigate the role played by PNPase in invasion gene expression.

# 1.4 Aims of the study

The aim of this study is to explore the interconnections between the transcriptional and post-transcriptional regulatory elements of *S*. Typhimurium by investigating the protein factors that regulate the expression of *S*. Typhimurium sRNAs.

It has become clear that control of gene expression at the level of transcription initiation is not the only method of modulating cellular genetic output, and that sRNA post-transcriptional regulators are also an important contributors to bacterial regulatory networks. Both transcriptional and post-transcriptional control points are integrated to ensure the correct spatial and temporal environment for gene expression to avoid incurring fitness costs. This is especially important during pathogenesis, when *S*. Typhimurium encounters a series of stressful environments and must out-compete the host microbiota to establish successful infection. *S*. Typhimurium is a well-adapted pathogen for host cell invasion, intracellular survival and replication, as well as the dissemination and establishment of systemic infection, as a result of its carefully co-ordinated programme of gene expression. Understanding the conditions when sRNAs

RNAs. A recent study from our laboratory addressed this question on a large-scale by using RNA-seq to profile the transcriptome of wild-type *S*. Typhimurium 4/74 grown under 22 infection-relevant environmental conditions (Kröger et al., 2013). The study by Kröger et al identified 280 *S*. Typhimurium sRNAs which are expressed under at least one infection-relevant condition. Many of the novel sRNAs are specific to the *Salmonella* genus, and indeed to *S. enterica*, suggesting that the sRNA genes were acquired to provide a fitness benefit to *S. enterica*, as the pathogenic serovars diverged from non-pathogenic *E. coli* and *S. bongori*.

Our aim in this study was to advance our knowledge about the 280 *S*. Typhimurium sRNAs by investigating the protein factors that control their expression on a transcriptional and post-transcriptional level. We chose 17 regulators that play an important role in the regulation of *S*. Typhimurium virulence gene expression, which comprise 3 σ-factors, 5 TCSs, 7 TFs involved in activation and repression of gene expression and 2 post-transcriptional regulators. The structural gene of each regulator was deleted and the mutant strain was grown under infection-relevant conditions known to activate the regulon of that protein. Wild-type *S*. Typhimurium 4/74 was used as a comparator for each mutant strain under each environmental condition. The transcriptome of each mutant and wild-type comparator was determined by RNA-seq to determine how expression of each sRNA gene may be regulated, either directly or indirectly. Finally, two *Salmonella*-specific sRNAs were chosen for detailed mechanistic analysis to study the mechanism involved in regulating their expression and the role played by the sRNA within the regulon of their cognate TF.

This study has expanded the transcriptional networks of some of the most important virulence-associated regulatory proteins of *S*. Typhimurium. In addition, the large-scale nature of the study has provided an overview of the overlap and integration between the transcriptional and post-transcriptional networks, or mixed regulatory networks, within the cell, in a manner that would not be possible through investigation of single components of the regulatory networks. Our hope is that the work presented in this thesis will provide a valuable community resource, upon publication, and that many putative regulatory interactions presented in this work will be explored and validated by other researchers in the future.

# Chapter 2 Materials & Methods

# 2.1 Chemicals and reagents

Unless otherwise stated, all chemicals and reagents were purchased from Becton Dickinson, Sigma-Aldrich, Roche, Qiagen, Chembio, New England Biolabs, Thermo Scientific, Ambion, Agilent, Bioline, Promega, Invitrogen and Millipore.

# 2.2 General microbiological techniques

#### 2.2.1 Media

Recipes for all liquid media are shown in Table 2.1. All media were made with water purified by de-ionising and filtration to a standard of 18.2 M $\Omega$ /cm (dH<sub>2</sub>0). All media were sterilised by autoclaving or vacuum-driven filtration through 0.22  $\mu$ M Steritop<sup>TM</sup> filter units.

Lennox agar was prepared by addition of 15 g/L Bacto agar to Lennox broth (Table 2.1) (Lennox, 1955) prior to autoclaving. Before use, agar was melted in a steamer and allowed to cool to 50°C before addition of antibiotics and pouring of agar plates.

#### 2.2.2 Culture conditions

Cells were routinely cultured in Lennox broth. Individual colonies were used to set up 4 mL overnight cultures in a glass test tube and these cultures were incubated at 37°C with agitation at 220 rpm in an Innova 40 air-incubator (New Brunswick Scientific) for 16-18 hours. Unless otherwise stated, overnight cultures were sub-inoculated 1:1,000 in 25 mL of pre-warmed Lennox broth in a 250 mL Erlenmeyer flask with appropriate antibiotics and grown with agitation at 220 rpm at 37°C in an Innova 3100 water-bath shaker (New Brunswick Scientific).

For growth in minimal media, overnight cultures were set up in Lennox broth as previously described. One mL of the overnight culture was harvested by centrifugation at 13,000 rpm at room temperature in an Eppendorf Centrifuge 5424. The cells were washed 3 times in pre-warmed minimal media and sub-inoculated 1:500 in 25 mL of minimal media in a 250 mL flask with appropriate antibiotics and grown with agitation at 220 rpm at 37°C in an Innova 3100 water-bath shaker. Optical density readings were

taken using a Biomate 3S spectrophotometer (Thermoscientific) at a wavelength of 600 nm.

When viable cell counts were required, 10-fold serial dilutions of a culture were carried out in sterile PBS. Typically, 100  $\mu$ L of dilutions of a factor of 10<sup>7</sup> were plated in duplicate and incubated at 37°C overnight. Cell numbers were calculated as colony forming units per ml (CFU/mL).

#### 2.2.3 Ectopic expression analysis

For ectopic expression of proteins from the arabinose-inducible  $P_{BAD}$  promoter, overnight cultures of strains carrying the pBAD plasmid containing the cloned gene and an empty pBAD vector were set up as previously described. Overnight cultures were diluted in 25 mL of the appropriate media in a 250 mL flask and grown to the desired  $OD_{600}$ . Glucose was occasionally used at a final concentration of 0.2% (where indicated) to repress expression from the  $P_{BAD}$  promoter (Guzman *et al.*, 1995). Cultures were split into two 250 mL flasks and L-arabinose was added at a final concentration of 0.2% to one flask, to induce expression from the  $P_{BAD}$  promoter, while the second culture was left un-induced. Induction proceeded for the desired length of time and cells were harvested for RNA isolation or preparation of whole cell lysates.

#### 2.2.4 Motility assay

Plate swimming assays were used to determine motility phenotypes. Semi-solid Lennox agar plates (Lennox broth, 0.3% agar) were prepared. The plates were divided into sections so swimming of each strain could be directly compared on the same plate. Each section was inoculated with 1  $\mu$ L of overnight culture, using a micropipette. The swimming plates were incubated at 37°C for 4 hours. Swimming diameters were then measured. Typically, every strain was assayed in triplicate in a minimum of 2 independent experiments.

#### 2.2.5 Antibiotics

Antibiotic stocks were made in  $dH_2O$  and filtered through a 0.22  $\mu$ M filter, or in 100% ethanol or in dimethyl sulfoxide (DMSO). Aliquots were stored at -20°C and thawed on ice before use. Antibiotics were used at the following final concentrations: Ampicillin (Amp) 100  $\mu$ g/mL (dH<sub>2</sub>O), Kanamycin (Kan) 50  $\mu$ g/mL (dH<sub>2</sub>O), Chloramphenicol (Cm) 35  $\mu$ g/mL (ethanol), Rifampicin (Rif) 150  $\mu$ g/mL (DMSO).

## 2.2.6 Maintenance of bacterial stocks

All strains and plasmids used in this study are listed in Table 2.2. For long term storage of bacteria, overnight cultures were mixed with a final concentration of 30% sterile glycerol and frozen at -80°C. When required, bacteria were streaked on Lennox agar containing appropriate antibiotics and were grown overnight at 37°C prior to sub-inoculation in liquid cultures. Bacteria were maintained on agar plates at 4°C for a maximum of one week.

Table 2.1 Liquid media recipes

Medium	Reagent	Concentration
Lennox	Bacto-Tryptone	10 g/L
	Sodium Chloride	5 g/L
	Bacto-Yeast extract	5 g/L
PCN (SPI2-inducing)	MES (pH 5.8)	80 mM
	Tricine	4 mM
	FeCl <sub>3</sub>	100 μΜ
	$K_2SO_4$	376 μΜ
	NaCl	50 mM
	K <sub>2</sub> HPO <sub>4</sub> /KH <sub>2</sub> PO <sub>4</sub> pH 5.8	0.4 mM
	glucose	0.4%
	NH <sub>4</sub> Cl	15 mM
	$MgSO_4$	1 mM
	CaCl <sub>2</sub>	0.01 mM
	$Na_2MoO_4$	10 nM
	$Na_2SeO_3$	10 nM
	$H_3BO_3$	4 nM
	$CoCl_2$	300 nM
	$CuSO_4$	100 nM
	$MnCl_2$	800 nM
	$ZnSO_4$	1 nM
Low Mg <sup>2+</sup> PCN	PCN medium w/o MgSO <sub>4</sub>	
	${ m MgSO_4}$	$10  \mu M$
MMA	$K_2HPO_4$	60 mM
	$\mathrm{KH_{2}PO_{4}}$	33 mM
	$(NH_4)_2SO_4$	7 mM
	dihydrate sodium citrate	1.7 mM
	${ m MgSO_4}$	1 mM
	Glycerol	0.4%

Table~2.2~Bacterial~strains~and~plasmids

Name	Genotype	Original Source/Reference
Bacterial strain	ns	
JH3676	wild-type S. Typhimurium 4/74. Str <sup>R</sup>	(Wray & Sojka, 1978)
JH3630	S. Typhimurium 4/74 ΔrpoE::frt	(Papenfort et al., 2006) <sup>a</sup>
JH3632	S. Typhimurium 4/74 Δfur::frt	JH3305 <sup>b</sup>
JH3635	S. Typhimurium 4/74 ΔhilD::frt	(Bustamante et al., 2008)
JH3636	S. Typhimurium 4/74 ΔslyA::kan	This Study
JH3637	S. Typhimurium 4/74 ΔphoB/R::kan	JH3312
JH3641	S. Typhimurium 4/74 Δfnr::frt	JH3307
JH3643	S. Typhimurium 4/74 ΔfliZ::frt	This Study
JH3646	S. Typhimurium 4/74 ΔrpoN::frt	JVS-1229 <sup>c</sup>
JH3649	S. Typhimurium 4/74 Δpnp::frt	JVS-0927
JH3652	S. Typhimurium 4/74 \(\Delta ompR/envZ::frt\)	(Cameron & Dorman, 2012)
JH3653	S.Typhimurium 4/74 \(\Delta dam::frt\)	(Balbontin et al., 2006)
JH3657	S. Typhimurium 4/74 \(\Delta ssrA/B::frt\)	This Study
JH3675	S. Typhimurium 4/74 \(\Delta ssrA::frt\)	This Study

Table 2.2 Bacterial strains and plasmids

JH3733	S. Typhimurium 4/74 \(\Delta ssr B:: frt\)	This Study
JH3660	S. Typhimurium 4/74 ΔphoP/Q::frt	This Study
JH3662	S. Typhimurium 4/74 ΔphoP::kan	JVS-8781
JH3763	S. Typhimurium 4/74 ΔphoP::frt	JVS-8781
JH3673	S. Typhimurium 4/74 ΔbarA/sirA-432::kan	This Study
JH3674	S. Typhimurium 4/74 ΔrpoS::kan	(Lucchini et al., 2009)
JH3584	S. Typhimurium 4/74 Δhfq::kan	JVS-00225
JH3765	S. Typhimurium 4/74 \(\Delta hilA::frt\)	JVS-1195
JH3766	S. Typhimurium 4/74 \(\Delta hilC::cat\)	Provided by W. D. Hardt
JH3767	S. Typhimurium 4/74 ΔhilE::cat	~ ADDIN EN.CITE (Sturm et al., 2011)
JH3768	S. Typhimurium 4/74 \(\Delta rtsA/B::kan\)	Constructed by Meabh Henry-Bezy
JH3769	S. Typhimurium 4/74 \(\Delta sprB\) \(hycA\)-STM2854 IGR::kan	Original strain: provided by J. T Wade. <i>kan</i> insertion and co-transduction: This Study
JH3774	S. Typhimurium SL1344 ΔinvF::kan	(Bustamante et al., 2008)
JH3775	S. Typhimurium SL1344 ΔSPI1::frt	(Paetzold et al., 2007)
JH3771	S. Typhimurium 4/74 \(\Delta\)STnc520::frt	Constructed by Meabh Henry-Bezy

Table 2.2 Bacterial strains and plasmids

JH3773	S. Typhimurium 4/74 ΔSTnc1480::frt	This Study
JH3774	S. Typhimurium 4/74 hns-1::kan	Mutant construction (Falconi <i>et al.</i> , 1991), Description (Hinton <i>et al.</i> , 1992)
JH3775	S. Typhimurium 4/74 phoP-3xFLAG (C-terminal)	This Study
JH3776	S. Typhimurium 4/74 slyA-3xFLAG (C-terminal)	This Study
JH3777	S. Typhimurium 4/74 hns-3xFLAG::kan (C-terminal)	(Dillon et al., 2010)
JH3778	S. Typhimurium 4/74 sprB-3xFLAG hycA-STM2854 IGR::kan	Original strain: provided by J. T Wade. kan insertion and co-transduction: This Study
JH3780	S. Typhimurium 4/74 hilA-3xFLAG::kan	This Study
JH3784	S. Typhimurium 4/74 ΔSTnc1480 hilA-3xFLAG::kan	This Study
JH3781	S. Typhimurium 4/74 fliC-3xFLAG::kan	JH3057
JH3785	S. Typhimurium 4/74 ΔSTnc1480 fliC-3xFLAG::kan	JH3057
JH3782	S. Typhimurium 4/74 ΔslyA hns-1::kan	This Study
JH3783	S. Typhimurium 4/74 ΔphoP hns-1::kan	This Study
TOP10	E. coli TOP10 F- mcrA $\Delta$ (mrr-hsdRMS-mcrBC) φ80lacZ $\Delta$ M15 $\Delta$ lacX74 nupG recA1 araD139 $\Delta$ (ara-leu)7697 galE15 galK16 rpsL(Str <sup>R</sup> ) endA1 $\lambda$ -	Invitrogen

Table 2.2 Bacterial strains and plasmids

Plasmids		
pKD3	Template plasmid for gene deletion; Amp <sup>R</sup> , FRT-flanked cat gene	(Datsenko & Wanner, 2000)
pKD4	Template plasmid for gene deletion; Amp <sup>R</sup> , FRT-flanked kan gene	(Datsenko & Wanner, 2000)
pKD46	Red helper plasmid; repA101 (ts), pBAD-γ-β-exo, araC <sup>+</sup> , Amp <sup>R</sup>	(Datsenko & Wanner, 2000)
pCP20	Plasmid for temperature sensitive FLP synthesis; ts-rep, FLP, Cm <sup>R</sup> , Amp <sup>R</sup>	(Cherepanov & Wackernagel, 1995)
pSUB11	Template plasmid for epitope tagging; Amp <sup>R</sup> , FRT-flanked 3xFLAG tag and kan	(Uzzau et al., 2001)
pBAD-myc-hisA	gene pBR322 origin, araC <sup>+</sup> , araBAD, C-terminal myc and polyhistidine tags, Amp <sup>R</sup>	Invitrogen
pBAD-phoP	phoP gene cloned in MCS of pBAD-myc-hisA	This study
pBAD-slyA	slyA gene cloned in MCS of pBAD-myc-hisA	This Study
pBAD- <i>hilD</i>	hilD gene cloned in MCS of pBAD-myc-hisA	(Sittka et al., 2008)
pBAD-STnc520	STnc520 gene cloned upstream of initiation codon of pBAD-myc-hisA	This Study
pBAD-STnc1480	STnc1480 gene cloned upstream of initiation codon of pBAD-myc-hisA	Constructed by Kristian Händler
pJV300	Control plasmid for 2-plasmid validation; ColE1, P <sub>LlacO</sub> promoter drives	(Sittka et al., 2007)
pP <sub>L</sub> -STnc520	transcription of 50 nt nonsense RNA derived from <i>rrnB</i> terminator, Amp <sup>R</sup> STnc520 gene constitutively transcribed P <sub>LlacO</sub> promoter on P <sub>PL</sub> vector, Amp <sup>R</sup>	This Study
pP <sub>L</sub> -STnc1480	$STnc1480$ gene constitutively transcribed $P_{LlacO}$ promoter on $P_{PL}$ vector, $Amp^R$	Constructed by Kristian Händler
pXG10 (sf) <sup>d</sup>	Plasmid for translational fusions of target mRNAs to GFP for 2-plasmid validation; pSC101* origin, P <sub>LtetO-1</sub> promoter, superfolder (sf) GFP, Cm <sup>R</sup>	(Corcoran et al., 2012)

<sup>&</sup>lt;sup>a</sup> Mutations transferred from original strains to 4/74 background by P22 transduction

<sup>&</sup>lt;sup>c</sup> Vogel lab strain collection

<sup>&</sup>lt;sup>b</sup> Hinton lab strain collection

<sup>&</sup>lt;sup>d</sup> pXG10 (sf) fusions used in this study listed in Appendix VII

# 2.3 General molecular techniques

#### 2.3.1 Preparation of chromosomal S. Typhimurium 4/74 DNA for use in PCR

Overnight cultures were set up as previously described. 400  $\mu$ L of the overnight culture were harvested by centrifugation at 13,000 rpm at room temperature for 1 minute. The pellet was re-suspended in 200  $\mu$ L nuclease free water and boiled at 100°C for 5 minutes, followed by vigorous vortexing to ensure full cell lysis. Cell debris was removed by centrifuging at 13,000 rpm for 5 minutes. The supernatant was transferred to a new 1.5 mL tube and 1 volume of chloroform was added, followed by vigorous vortexing for 30 seconds. The organic and aqueous phases of the mixture were separated by centrifugation at 13,000 rpm for 10 minutes at 4°C. The aqueous DNA-containing layer was transferred to a new 1.5 mL tube and the concentration was adjusted to 100 ng/ $\mu$ L with nuclease free water.

#### 2.3.2 Polymerase chain reaction (PCR)

PCR amplifications were carried out in a Mastercycler ep gradient S thermocycler (Eppendorf). For routine amplifications, Taq DNA polymerase (MangoTaq<sup>TM</sup> supplied by Bioline or Taq purified from an E. coli strain carrying pUC18 containing the cloned coding region for the taq gene as previously described (Desai & Pfaffle, 1995)) was used. Both Taq enzymes were used with the 5X reaction buffer and MgCl<sub>2</sub> supplied with the MangoTaq<sup>TM</sup> enzyme. A mastermix containing a 1.3X reaction buffer, 4 mM MgCl<sub>2</sub>, 1.1 mM total dNTPs and nuclease free water was assembled and aliquots were stored at -20°C. In a typical 20 μL PCR reaction the mastermix was diluted 1.3-fold, primers were added to a final concentration of 1 pmol/µL and 1 unit of enzyme was used. One unit of MangoTaq<sup>TM</sup> enzyme is defined as the amount of enzyme that incorporates 10 nmoles of dNTPs in 30 minutes at 72°C. For applications requiring high fidelity amplification of a PCR product Velocity DNA polymerase (Bioline) was used. The Velocity enzyme possesses 5' to 3' and 3' to 5' proofreading exonuclease activity, with a low error rate of 4.4 x 10<sup>-7</sup>. The enzyme was used with the supplied 5X Hi-Fi reaction buffer, containing 10 mM MgCl<sub>2</sub>. dNTPs were used at a total final concentration 0.8 mM, primers were used at a concentration of 1 pmol/µL and 1 unit of enzyme was used per reaction.

Initial denaturation in the reaction was dependent on the template DNA used. For colony PCR initial denaturation proceeded for 10 minutes at 95°C and the time was reduced to 2 minutes when extracted and purified DNA was used as template. Initial denaturation was followed by 30 cycles of denaturation for 15 seconds at 95°C, annealing for 30 seconds at a temperature typically 4°C below the lowest melting temperature of the primer pair and extension for approximately 15-30 seconds per kilobase at 72°C. The cycling was followed by a final extension step for 7 minutes at 72°C. PCR products were held at 4°C until they were analysed.

#### 2.3.3 Agarose gel electrophoresis

DNA fragments were separated based on their size using agarose gel electrophoresis. Samples were mixed with tri-coloured loading dye (40% glycerol, 60 mM EDTA, 10 mM Tris-HCl pH 7.6, 0.03-0.25% bromophenol blue pH 8.0, 0.03-0.25% xylene cyanol FF, 0.15-0.25% orange G) to a final concentration of 1X loading dye. Typically 2% Tris Acetate Ethylenediaminetetraacetic acid (TAE) gels were run in 1X TAE buffer at a constant voltage of 120 V. After electrophoresis, agarose gels were stained for at least 1 hour in Redsafe<sup>TM</sup> nucleic acid stain and were visualised under ultra violet light in an AlphaImager 2200. The molecular size of electrophoresed DNA fragments was estimated by comparison with Hyperladder I or Hyperladder IV (Bioline).

#### 2.3.4 Restriction digestion

All restriction enzymes were purchased from New England Biolabs or Thermo Scientific and restriction digests were carried out according to manufacturers' protocols regarding reaction buffer, temperature, timing of reaction and enzyme inactivation for each individual restriction enzyme. Digested DNA fragments or plasmids were then analysed by agarose gel electrophoresis as previously described.

#### 2.3.5 Purification of PCR products

PCR products were typically purified to remove primers, enzymes, nucleotides and other impurities using a QIAquick PCR purification kit (Qiagen). A high salt binding buffer (PB buffer containing guanidine hydrochloride) was added to the sample and 10  $\mu$ L 3M sodium acetate pH 5.2 was added to reduce the pH to pH 7.5 or lower, as determined by pH indicator in the binding buffer. The sample was passed through a silica membrane-containing column by centrifugation at 13,000 rpm for 1 minute at room temperature. Nucleic acids bind to the silica membrane under high salt conditions and at a pH less than 7.5. The column-bound DNA was washed using an ethanol-based wash buffer and centrifugation at 13,000 rpm for 1 minute at room temperature twice. 30  $\mu$ L of nuclease free water was added and the column was left to settle for 1 minute at room temperature. The column was centrifuged at 13,000 rpm for 1 minute and the DNA was eluted from the column due to the reduction in salt concentration and more basic pH. DNA was stored at -20°C.

### 2.3.6 Extraction of DNA from agarose gels

The QIAquick gel extraction kit (Qiagen) was used to extract and purify DNA from agarose gels. TAE agarose gels were run and stained as previously described. The electrophoresed DNA fragments were visualised under UV light and the specific band of interest was excised from the gel using a clean scalpel. The DNA-containing gel was weighed and a volume, equal to 3 times the weight of the gel, of solubilisation and binding buffer (QG buffer) was added. The gel was incubated in solubilisation buffer at 50°C with occasional vortexing until the gel slice was fully dissolved. 3M sodium acetate pH 5.2 was added to adjust the pH of the sample to below 7.5 to increase efficiency of DNA binding to the silica membrane column. DNA binding, washing of the column and elution of DNA proceeded as before.

#### 2.3.7 Plasmid preparations

Plasmid DNA was extracted and purified using QIAprep Spin mini- or midi- prep kits (Qiagen). For routine plasmid preparation, mini-prep kits were used. 5 mL of an

overnight culture was harvested by centrifugation at 4,000 rpm for 10 minutes at room temperature. Bacterial pellets were re-suspended and lysed under alkaline conditions in buffers P1 (TE buffer pH 8.0, 100  $\mu$ g/mL RNaseA) and P2 (200 mM NaOH, 1% SDS) according to manufacturers' instructions. The cell lysate was neutralised by addition of the high salt containing binding buffer (N3 buffer containing guanidine hydrochloride and acetic acid) and cellular debris was pelleted by centrifugation at 13,000 rpm for 10 minutes at room temperature. The supernatant was applied to a silica membrane-containing spin column and the plasmid DNA was bound to the membrane (due to the high salt and reduced pH conditions) by centrifugation at 13,000 rpm for 1 minute at room temperature. The column was washed and plasmid DNA was eluted as previously described.

Midi-prep kits were used when larger amounts of plasmid were required. Starter cultures (4 mL) were inoculated and grown for at least 8 hours at 37°C with appropriate antibiotics. The starter culture was then sub-inoculated 1:1,000 into a larger culture volume according to manufacturers' recommendations on plasmid copy number. Cultures were grown for approximately 16 hours. The entire cultures were harvested by centrifugation at 13,000 rpm for 30 minutes at 4°C. Pellets were re-suspended and lysed as with the mini-prep kit but with larger volumes of each buffer. After addition of a high salt neutralisation buffer (buffer P3 containing potassium acetate), samples were incubated on ice for 15 minutes to aid in precipitation of cellular debris. Cellular debris was pelleted by centrifugation at 4,000 rpm for 120 minutes at 4°C. The supernatant was centrifuged for a further 90 minutes to ensure full removal of particles. A QIAGEN-tip 100 was equilibrated with buffer QBT (750 mM NaCl, 50 mM MOPS pH 7.0, 15% isopropanol, 0.15% Triton X-100) and allowed to empty by gravity flow. plasmid-containing supernatant was then added to the column and allowed to empty by gravity flow. The column-bound DNA was washed twice in buffer QC (1 M NaCl, 50 mM MOPS pH 7.0, 15% isopropanol). The plasmid DNA was eluted from the column with buffer QF (1.25 M NaCl, 50 mM Tris-Cl pH 8.0, 15% isopropanol). The plasmid DNA was precipitated from the eluate by adding 0.7 volumes of isopropanol and centrifugation at 4,000 rpm for 90 minutes at 4°C. The pellet was washed in 70% ethanol and allowed to air dry. The pellet was re-suspended in 50 µL nuclease free water and the DNA was stored at -20°C.

# 2.4 Preparation and transformation of competent bacteria

## 2.4.1 Preparation of electro-competent cells and electroporation (Dower et al., 1988)

Overnight S. Typhimurium cultures were diluted 1:100 in 25 mL Lennox broth with appropriate antibiotics and grown with agitation at 220 rpm to exponential phase (OD<sub>600</sub> 0.4-0.6). The culture was then split in half and the cells were harvested by centrifugation at 4,000 rpm at 4°C for 10 minutes in an Eppendorf centrifuge 5810R. Cells were washed in 10 mL ice-cold water and harvested by centrifugation at 4,000 rpm at 4°C for 10 minutes. Cells were re-suspended in 1 mL of ice cold water and moved to a 1.5 mL centrifuge tube. The cells were washed a further three times with 1 mL of ice-cold water and centrifugation at 13,000 rpm at 4°C for 5 minutes per wash. The electro-competent cells were finally re-suspended in 200 µL of ice-cold water. 40 µL of electro-competent cells were transformed with plasmids or linear DNA fragments using a GenePulser Xcell electroporator (Biorad) and 2 mm electroporation cuvettes at 2.5 kV. Negative controls containing electro-competent cells without added DNA were also electroporated. One mL of Lennox broth was added to the electroporated cells and the cells were recovered for 1 hour at 37°C with agitation at 220 rpm. A 100-150 μL aliquot of cells was plated on an appropriate antibiotic selection plate and plates were incubated overnight. remaining cells were incubated at room temperature overnight and were plated if there were no transformants after overnight incubation of the selection plates.

#### 2.4.2 Preparation and transformation of chemically competent bacteria

Overnight *E. coli* cultures were diluted 1:100 in 50 mL Lennox broth with appropriate antibiotics and grown with agitation at 220 rpm to exponential growth phase ( $OD_{600}$  0.4-0.8). Cells were harvested by centrifugation at 4000 rpm for 10 minutes at 4°C. The cells were re-suspended in 40 mL ice-cold 100 mM  $CaCl_2$  and incubated on ice for 20 minutes. Cells were harvested as before and re-suspended in 10 mL ice cold 100 mM  $CaCl_2$ . Glycerol was added to a final concentration of 10% and the cells were incubated on ice for a further 20 minutes. The chemically competent cells were stored at -80°C in 300  $\mu$ L aliquots.

An entire aliquot of chemically competent cells was used for each transformation. The DNA was added to the aliquot of chemically competent cells and the cells were incubated on ice for 20 minutes. The cells were heat-shocked for 90 seconds at 42°C, followed by incubation on ice for a further 2 minutes.  $800~\mu L$  of Lennox broth was added to each tube of transformed cells and the cells were allowed to recover at  $37^{\circ}C$  with agitation (220 rpm) for 1 hour. Negative controls containing chemically competent cells without added DNA were treated in the same way. All cells were harvested by centrifugation in a table-top centrifuge at room temperature for 5 minutes at 13,000~rpm. The cell pellets were resuspended in  $100~\mu L$  of Lennox broth and were plated on an appropriate antibiotic selection plate and plates were incubated overnight.

# 2.5 Genetic manipulations

#### 2.5.1 Generation of deletion mutants in S. Typhimurium 4/74

Gene deletion mutants were generated using the Datsenko and Wanner method (Datsenko & Wanner, 2000). Kanamycin or Chloramphenicol resistance cassettes were amplified from the pKD4 or pKD3 plasmids respectively. Amplification was carried out in 50 μL reactions using the proof-reading DNA polymerase Velocity. Primers used to amplify plasmid-encoded resistance cassettes also contained sequences with 50 base pair homology to regions flanking the gene of interest (as shown in Appendix IX). Approximately 1 ng of either plasmid preparation was used as template for PCR. PCR cycling conditions were as follows:

Initial Denaturation	98°C	30 seconds
Denaturation	98°C	10 seconds
Annealing	54°C	1 minute $\rightarrow$ 35 cycles
Extension	72°C	1 minute
Final Extension	72°C	10 minutes

PCR products were purified using a QIAquick PCR purification kit according to the manufacturer's instructions. Purified PCR products were visualised using UV light

following agarose gel electrophoresis in 1X Tris-Acetate-EDTA buffer and staining in Redsafe<sup>Tm</sup> nucleic acid stain.

4/74 containing the  $\lambda$ -Red-encoding pKD46 plasmid was supplemented with Ampicillin and 1-10 mM L-arabinose and grown overnight at 30°C with agitation at 220 rpm. Overnight cultures were diluted 1:100 in 25 mL of Lennox broth, supplemented with Ampicillin and 1-10 mM L-arabinose. Cells were made electro-competent as previously described. Electro-competent 4/74 pKD46 cells were transformed with 200 ng, 400 ng and 600 ng of purified DNA fragment containing the antibiotic resistance cassette and sequences homologous to the flanking regions of the gene of interest. The antibiotic resistance cassette was integrated into the bacterial chromosome, in place of the gene of interest, by homologous recombination mediated by the  $\lambda$ -Red recombinase system (encoded by  $\gamma$ ,  $\beta$ , exo genes). Mutants were selected with appropriate antibiotic plates and overnight growth at 37°C. Mutations were moved into clean genetic backgrounds by P22 transduction, and confirmed by PCR and DNA sequencing (Source Biosciences, Dublin).

#### 2.5.2 Epitope tagging

Genes were tagged in their chromosomal location using an adapted  $\lambda$  Red recombineering protocol as previously described (Uzzau et al., 2001). Briefly, 3 FLAG tags preceding the Kanamycin resistance cassette were amplified from the pSUB11 plasmid using primers with 40 base pair homology to the 3' end of the gene to be tagged as shown in Appendix IX. Amplification was carried out in 50  $\mu$ L high fidelity reactions as previously described. The PCR products were purified and analysed by agarose gel electrophoresis. This linear DNA fragment was integrated into the bacterial chromosome by transformation of a strain carrying the pKD46 plasmid by homologous recombination as previously described. Transformants carrying the FLAG-tagged gene were selected for on Kanamycin plates and the presence of the tag was confirmed by colony PCR, sequencing and western immunoblotting. Tagged genes were moved into clean genetic backgrounds by P22 transduction, and confirmed by PCR and DNA sequencing.

#### 2.5.3 P22 phage transduction

To generate P22 phage lysates, overnight cultures of the donor strain were diluted 1:1,000 with appropriate antibiotics in 15 mL of Lennox broth and grown to an  $OD_{600}$  0.1. The cells were incubated with 20  $\mu$ L of P22 HT 105/1 *int-201* lysate generated from wild-type 4/74 cells. Phage infection proceeded for 4 hours. Remaining bacterial cells were killed by addition of 500  $\mu$ L chloroform and incubation at room temperature for 10 minutes. Bacterial cells and cellular debris were removed by centrifugation at 4,000 rpm for 20 minutes at room temperature. The phage-containing supernatant was filtered through a Millex 0.22  $\mu$ M filter (Millipore) and stored at 4°C.

The recipient strain was grown in 4 mL of Lennox broth in a test tube at 37°C with shaking at 220 rpm overnight. 100 μL of the recipient strain was incubated with 100 μL of neat, 10<sup>-1</sup> and 10<sup>-2</sup> dilutions of the donor P22 lysate at 37°C for 1 hour without shaking. "Donor only" and "recipient only" negative controls were also incubated at 37°C for 1 hour without shaking. The entire transduction mixtures were plated on appropriate antibiotic selection plates and incubated overnight at 37°C. Transductants were chosen from the plate with the highest phage dilution and grown in 4 mL of Lennox broth with appropriate antibiotics and 10 mM Ethylene glycol-bis (2-aminoethylether)-N, N, N', N'-tetraacetic acid (EGTA). These cultures were then streaked on Green agar indicator plates (Lennox agar, green agar dye stock consisting of 62.5 mg/mL mordant orange and 6.5 mg/mL aniline blue) with appropriate antibiotics and incubated overnight at 37°C. Green agar plates are used to screen against unstable pseudo-lysogens, which produce dark green colonies on the plates due to bacterial lysis causing a reduction in pH. Pale colonies were picked from the Green agar plate and screened a second time on a Green agar plate. One pale colony was then chosen to be stored as glycerol stock at -80°C.

#### 2.5.4 Removal of antibiotic resistance cassettes

To ensure antibiotic resistance cassettes would not affect transcription of downstream genes, the resistance genes were removed from mutant strains using the pCP20 plasmid (Datsenko & Wanner, 2000, Cherepanov & Wackernagel, 1995). The pCP20 plasmid encodes the yeast FLP recombinase and is temperature sensitive for replication and induction of FLP expression. FLP recognises the FLP recombinase target sites (FRT

sites) which flank the antibiotic resistance determinant in the mutant strain. FLP mediates site-specific recombination between the two FRT sites resulting in excision of the antibiotic resistance gene, leaving behind an 82-85 nt scar sequence. Antibiotic resistant mutants or tagged strains were grown to mid-log phase and cells were made electrocompetent as previously described. Electroporation was used to transform cells with 500 ng of pCP20 plasmid. Cells were allowed to recover for 1 hour at 30°C with aeration. 150 μL of transformation mixes were plated on chloramphenicol plates (or on ampicillin plates if a chloramphenicol resistance gene was to be removed) and plates were incubated at 30°C overnight. Transformants were passaged at 37°C without antibiotics to cure the strain of the pCP20 plasmid. Loss of the antibiotic resistance cassette and the pCP20 plasmid were screened for using appropriate antibiotic selection plates and incubation at permissive temperatures.

#### 2.5.5 Sequence and Ligation Independent Cloning (SLIC)

Sequence and ligation independent cloning (SLIC) was carried out as adapted from (Li & Elledge, 2007). The insert and vector backbone were amplified using Velocity high fidelity DNA polymerase and the following cycling conditions respectively:

Initial Denaturation	98°C	1 minute
Denaturation	98°C	20 seconds
Annealing	55°C	$30 \text{ seconds}$ $\rightarrow$ $30 \text{ cycles}$
Extension	72°C	1 minute
Final Extension	72°C	5 minutes
Storage	4°C	
Initial Denaturation	98°C	2 minutes
Denaturation	98°C	30 seconds
Annealing	50-65°C (gradient)	$30 \text{ seconds/}^{\circ}\text{C} > 35 \text{ cycles}$
Extension	72°C	3.5 minutes
Final Extension	72°C	5 minutes
Storage	4°C	

Primers to amplify the insert were designed to contain approximately 20 base pairs of homology to the cloning site of the vector as shown in Appendix IX. Wild-type chromosomal DNA and plasmid DNA were used as templates to amplify the insert DNA and vector backbone respectively. Insert DNA and vector backbone were purified using the QIAquick PCR purification kit according to the manufacturer's instructions. 1 µg of template plasmid backbone was digested using DpnI restriction enzyme for 1 hour at 37°C, followed by heat inactivation of the enzyme at 80°C for 20 minutes. 1 μg of DpnI-digested vector backbone and 1µg of insert DNA were treated with 5 units of T4 DNA polymerase in the absence of deoxynucleotide triphosphates to generate single strand overhangs. T4 DNA polymerase treatment proceeded at 23°C for 30 minutes in the presence of 5 mM DTT, 200 mM Urea, 1X BSA, and 1X reaction buffer (67 mM Tris-HCl pH8.8, 6.6 mM MgCl<sub>2</sub>, 1mM DTT, 16.8 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>). The reaction was stopped by addition of 25 mM EDTA and incubation at 75°C for 20 minutes. 100 ng of T4 DNA polymerase treated vector was mixed with an equal volume of T4 DNA polymerase treated insert in a final volume of 10 μL. Samples were incubated at 65°C for 10 minutes followed by "touch-down" annealing, during which the incubation temperature was reduced from 65°C to 25°C in 1°C decrements and samples were held for 1 minute at each temperature. Annealed vector and insert mixtures were then transformed into chemically competent E. coli TOP10 cells as previously described. Positive clones were selected with appropriate antibiotic selection plates and overnight incubation at 37°C. Clones were screened by colony PCR using plasmid-specific and gene-specific primers. The inserted DNA was then sequenced.

# 2.6 Analysis of RNA

# 2.6.1 Extraction of RNA from S. Typhimurium 4/74

Overnight cultures were diluted in 25mL of pre-warmed media in a 250 mL flask with appropriate antibiotics and grown to the desired  $OD_{600}$  as previously described. Four or 5  $OD_{600}$  units were removed from the culture, and cellular transcription was stopped using 0.4X culture volume of a 5% phenol (pH 4.3) 95% ethanol "stop" solution (Sigma P4557 and E7023, respectively). Transcription was stopped and RNA was stabilised on ice in stop solution for at least 30 minutes before cells were harvested at 4,000 rpm for 10

minutes at 4°C. At this point pellets were either stored at -80°C, or RNA was immediately extracted. To isolate RNA, pellets were re-suspended in 1 mL of TRIzol<sup>TM</sup> 400 µL of chloroform was added and the samples were Reagent (Invitrogen). immediately and thoroughly mixed by inversion. Samples were moved to a Phase-lock tube (5 Prime) and the aqueous and organic phases were separated by centrifugation at 13,000 rpm for 15 minutes at room temperature in a table top centrifuge. The aqueous phase was moved to a new 1.5 mL tube and the RNA was precipitated using isopropanol for 30 minutes at room temperature followed by centrifugation at 13,000 rpm for 30 minutes at room temperature. The RNA pellet was rinsed with 70% ethanol followed by centrifugation at 13,000 rpm for 10 minutes at room temperature. The RNA pellet was air-dried for 15 minutes and re-suspended in DEPC-treated water at 65°C with shaking at 900 rpm on a Thriller thermoshaker (Peqlab) for 5 minutes with occasional vortexing. RNA was kept on ice whenever possible and RNA was stored at -80°C. concentration was measured using a nanodrop ND-1000 Spectrophotometer and RNA quality was assessed using a 2100 Bioanalyser (Agilent).

#### 2.6.2 DNase I digestion

RNA to be used for cDNA library preparations was treated with 10 units of DNase I (Fermentas) to remove any DNA present in the sample. DNase I digestion proceeded for 40 minutes at 37°C in the presence of 10 mM Tris-HCl pH 7.5, 2.5 mM MgCl<sub>2</sub> and 0.1 mM CaCl<sub>2</sub> in a final volume of 100 µL. The reaction was stopped by addition of addition of Phenol-Chloroform:Isoamyl Alcohol 25:24:1. The samples were mixed immediately and thoroughly by inversion and the aqueous and organic phases were separated by centrifugation at 13,000 rpm at 15°C for 12 minutes. The aqueous phase was moved to a new 1.5 mL tube and the DNase I treated RNA was precipitated by addition of 2.5X volume of a 30:1 mixture of 100% Ethanol and 3M Sodium Acetate (pH 6.5). Precipitation proceeded at -20°C for at least 1 hour. The RNA was pelleted by centrifugation at 13,000 rpm for 30 minutes at 4°C. The RNA pellet was rinsed with 70% ethanol by centrifugation at 13,000 rpm for 10 minutes at 4°C. The RNA pellet was airdried for 20 minutes and re-suspended in DEPC-treated water at 65°C with shaking at 900 rpm on a Thriller thermoshaker for 5 minutes with occasional vortexing. RNA

concentration was measured using a nanodrop ND-1000 spectrophotometer and RNA quality was assessed using a Bioanalyser. RNA was stored at -80°C.

#### 2.6.3 Reverse transcription

First-strand cDNA synthesis was carried out using the GoScript<sup>TM</sup> Reverse Transcription system (Promega). 400 ng of total DNase I digested RNA was combined with 0.5  $\mu$ g of random primers in a final volume of 5  $\mu$ L. The mixture was incubated at 70°C for 5 minutes and quickly chilled on ice for 5 minutes. The reverse transcription mastermix was prepared on ice and contained a final concentration of 1X GoScript<sup>TM</sup> reaction buffer, 3 mM MgCl<sub>2</sub>, 0.5 mM of each dNTP, 20 U recombinant RNasin Ribonuclease inhibitor, 160 U GoScript<sup>TM</sup> Reverse Transcriptase (RT). The reverse transcription mastermix was added to the RNA-primer mix to a final volume of 20  $\mu$ L on ice. "No reverse transcriptase" and "no template" controls were included to confirm the absence of contaminating DNA in the RNA samples and reagents respectively. Annealing of the random primers to the RNA templates proceeded at 25°C for 5 minutes, followed by elongation of the cDNA products at 42°C for 1 hour. The reverse transcriptase was inactivated by incubation at 70°C for 15 minutes. Following cDNA synthesis, 100  $\mu$ L of nuclease free water was added to each sample to make a final concentration of 3.33 ng/ $\mu$ L and cDNA samples were stored at -20°C.

#### 2.6.4 Quantitative real-time PCR

Quantitative real-time PCR (qRT-PCR/qPCR) was used to precisely quantify gene expression. Roche FastStart Universal SYBR Green mastermix with Rox (FastStart Taq DNA polymerase, reaction buffer, dNTPs, SYBR Green I, reference dye Rox) was used at a final concentration of 1X and primers were used at a concentration of 0.3  $\mu$ M in each 20  $\mu$ L qPCR reaction. Standard curves containing 5 serial 10-fold dilutions of wild-type 4/74 (WT) genomic DNA (prepared as described in section 2.3.1) were generated. Dilutions were not stored and were prepared fresh every time. 8  $\mu$ L of each dilution was used in each 20  $\mu$ L qPCR reaction and the standard curve was amplified separately using each primer set. cDNA templates, generated by reverse transcription (as previously described), were typically diluted 5-fold in nuclease-free water to a final concentration of

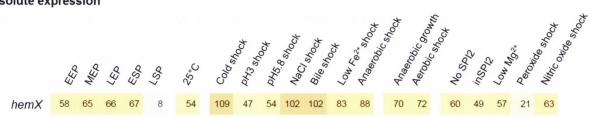
0.67 ng/ $\mu$ L, and 8  $\mu$ L (5.36 ng) of this dilution was used in each 20  $\mu$ L qPCR reaction. qPCR runs were carried out in MicroAmp Fast Optical 96-well reaction plates (Applied Biosystems) using an ABI StepOnePlus<sup>TM</sup> real-time PCR machine. Standard cycling conditions were as follows:

Hot Start step	95°C	10 minutes	
Denaturation	95°C	15 seconds	101
Annealing	60°C	1 minute	40 cycles

Melt curves (15 seconds at 95°C, 1 minute at 60°C, 0.3°C increments from 60°C to 95°C with a 1 minute hold at each temperature, 95°C for 15 seconds) were included after amplification for every new primer set to aid in ensuring the production of a specific PCR product. Analysis was done using the StepOne<sup>TM</sup> software. PCR products were quantified relative to the standard curve and were internally normalised to a reference gene. The gene *hemX*, which encodes a putative uroporphyrinogen III C-methyltransferase was typically used as internal reference gene as it showed the least variation in its expression levels across 20 environmental conditions (Kröger et al., 2013) and 18 mutants (this study) as demonstrated by RNA-seq (Figure 2.1).

A.





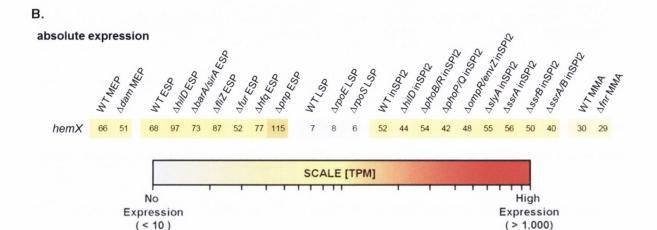


Figure 2.1 Absolute expression in TPM (transcripts per million) of the hemX gene

**A.** Screenshot from the SalCom website (Kröger et al., 2013) showing absolute expression of the *hemX* gene in wild-type 4/74 grown in 20 environmental conditions. **B.** Screenshot from the unpublished website SalCom version 3 (this study) showing absolute expression of the *hemX* gene in wild-type 4/74 and 18 isogenic mutants grown in 5 environmental conditions.

#### 2.6.5 Generating Digoxigenin-labelled riboprobes

Digoxigenin (Dig)-labelled riboprobes were generated by in vitro transcription using T7 RNA polymerase using a Dig Northern Starter kit (Roche). A linear DNA template incorporating the gene of interest and a T7 promoter sequence was first generated by PCR using primers as shown in Appendix IX and a high fidelity enzyme as previously described. The DNA template was purified by gel extraction according to the manufacturers' instructions and 100-200 ng of DNA template was used in each in vitro transcription reaction. Linear DNA template was combined with 1X labelling mix, which contains unlabelled nucleotides and DIG-11-UTP, 1X transcription buffer and 40 units of T7 polymerase in a final volume of 20 µL. Labelling transcription mixes were incubated at 42°C for 1 hour. 20 U of DNase I was added to remove template DNA and the reactions were incubated at 37°C for 15 minutes. The reaction was stopped by addition of 400 mM EDTA (pH 8.0) and labelled riboprobes were stored at -20°C. Labelling efficiency was determined by dot-blotting serial 10-fold dilutions of riboprobe onto positively charged nylon membrane and incubation of the membrane with a 1:10,000 dilution alkaline phosphatase-conjugated anti-digoxigenin Fab-fragments raised in sheep (as described in section 2.6.7). The alkaline phosphatase chemiluminescent substrate CDP-Star was used for signal detection. Signal was visualised following exposure in an ImageQuant LAS 4000 imager (GE Healthcare).

#### 2.6.6 Northern blotting

Total RNA was separated based on its size by electrophoresis through an 8.3 M Urea, 1X TBE (Tris Borate EDTA) polyacrylamide gel. For analysis of small RNAs by northern blotting, 7% polyacrylamide gels were used. Generally 5 µg of RNA was mixed with 2X RNA loading dye (1.25% xylene cyanol, 1.25% bromophenol blue, 450 mM EDTA pH 8.0, 0.03% SDS, 94% formamide) and samples were heated at 65°C for 10 minutes. Samples were run on the gel in 1X TBE running buffer at a constant voltage of 90 V. Gels containing the separated RNA, a positively charged nylon membrane and filter paper were immersed in 1X TBE and the RNA was transferred to the positively charged nylon membrane using the Biometra Fastblot B43 semi-dry blotting apparatus at a constant amplitude of 125 mA for 30 minutes at 4°C. RNA was UV-crosslinked to the membrane at 120 mJ (Peqlab cross-linker) for 2 minutes. The membrane was equilibrated in

hybridisation buffer for 1 hour at 62°C in pre-warmed DIG Easy Hyb solution in a rotating hybridisation oven. 5 µL (approximately 1.25 µg) of riboprobe was boiled at 95°C for 5 minutes and allowed to cool on ice for 5 minutes. The boiled riboprobe was added to the pre-hybridising solution and hybridisation proceeded at 62°C overnight. The membrane was washed twice for a total of 10 minutes in pre-heated (62°C) stringency wash buffer 1 (2X saline-sodium citrate (SSC) buffer, 0.1% SDS) at room temperature with rocking on a see-saw rocker (Stuart) at room temperature, followed by 2 washes for a total of 30 minutes with room temperature stringency wash buffer 2 (0.5X SSC buffer, 0.1% SDS) with rocking at room temperature. Non-specific sites on the membrane were blocked using 1X blocking buffer (casein-based blocking buffer supplied by Roche and diluted 10-fold in maleic acid buffer (0.1 M maleic acid, 0.15 M NaCl adjusted to pH 7.5 using NaOH pellets)) for 30 minutes at room temperature with rocking. phosphatase conjugated polyclonal anti-digoxigenin Fab-fragment was diluted 1:10,000 in 1X blocking buffer and immunological detection of the membrane proceeded for 30 minutes at room temperature with rocking. The membrane was then washed twice for a total of 30 minutes in wash buffer (maleic acid buffer, 0.3% Tween-20). The membrane was incubated for 5 minutes in detection buffer (0.1M Tris-HCl, 0.1M NaCl, pH 9.5) and CDP-star was used as the chemiluminescent substrate. Enzymatic de-phosphorylation of CDP-star by alkaline phosphatase results in light emission which was visualised using an ImageQuant Las 4000 Imager.

To determine if RNA samples were equally loaded on the gel, the membrane was reprobed for the 5S ribosomal RNA. Firstly, the membrane was quickly washed in sterile dH<sub>2</sub>0 and was then stripped using freshly prepared stripping buffer (50% formamide, 50 mM Tris-HCl pH 7.5, 5% SDS) at 80°C with rotation twice for 1 hour each. The membrane was then rinsed twice for 5 minutes each with 2X SSC at room temperature followed by pre-hybridisation at 68°C as before for 1 hour. The 5S riboprobe was diluted 1:1000 and boiled at 95°C for 5 minutes and allowed to cool on ice for 5 minutes. The diluted and boiled riboprobe was added to the pre-hybridising solution and hybridisation proceeded overnight at 68°C. Washing, blocking and immunological detection of the membrane were repeated as before.

#### 2.6.7 cDNA library preparation for RNA-seq

Strand-specific cDNA library preparation and high throughput cDNA sequencing (RNA-seq) of wild-type 4/74 and isogenic mutants, for investigation of the protein factors involved in regulating the expression of *S*. Typhimurium sRNAs, was performed on DNase I digested total RNA by Vertis Biotechnologie AG (Freising, Germany). Strains and conditions in which they were grown before RNA extraction are detailed in Table 2.3. RNA was not depleted for ribosomal RNA or enriched for small RNAs. RNA was fragmented by sonication and poly(A)-tails were added to each fragment by poly(A) polymerase. The 5' end of each fragment was de-phosphorylated using tobacco acid pyrophosphatase (TAP). An RNA adaptor, containing a 6-10 nucleotide bar-code was ligated to the 5' end of each fragment to allow for identification of each RNA fragment after sequencing. First strand cDNA synthesis was performed using oligo (dT) priming and Moloney murine leukaemia virus reverse transcriptase (M-MLV RT). The resulting cDNA was amplified by PCR to approximately 20-30 ng/µL using a high fidelity DNA polymerase. cDNA was purified using the Agencourt AMPure XP kit (Beckman Coulter Genomics) and analysed by capillary electrophoresis.

Strand-specific cDNA library preparation and high-throughput RNA-seq analysis of three samples, used for a transcriptomics-based sRNA target-hunt, was performed on DNase I digested total RNA by the Centre of Genomic Research (CGR), University of Liverpool. The strains and conditions under which they were grown before RNA was extracted are detailed in section 4.2.11. The RNA was first depleted for ribosomal RNA using the Ribo-Zero<sup>TM</sup> kit and strand-specific cDNA libraries were generated using the ScriptSeq<sup>TM</sup> kit. This library preparation method also involves the tagging of fragmented RNA with adaptor molecules for directional sequencing.

#### 2.6.8 RNA-seq

cDNA prepared by Vertis Biotechnologie AG was sequenced on an Illumina HiSeq 2000 platform. Firstly, cDNA templates are immobilised on a flow cell surface and are amplified by bridge amplification to create clusters of identical copies of single template molecules. Illumina sequencing uses "sequencing by synthesis" (SBS) technology, involving the incorporation of dNTPs reversibly labelled with fluorescent terminators.

The labels serve as terminators for polymerase so each dNTP fluoresces and is imaged to identify the individual nucleotide that is incorporated into the nucleic acid chain during each sequencing cycle. The label is then enzymatically cleaved to allow addition of the next nucleotide. In total, 3 individual sequence runs were carried out to investigate the protein factors involved in regulating expression of *S*. Typhimurium sRNAs. In the first run, 4 flow cell lanes were used containing between 7 and 9 samples per lane. This resulted in 100 million reads of 50 bp in length per flow cell lane. Three lanes were used in the second sequencing run containing between 5 and 7 samples per lane. Read lengths were increased to 100 bp in this run. Two lanes were used in the third sequencing run containing between 5 and 7 samples per lane. Mutants were always sequenced in the same lane as their wild-type comparator.

cDNA prepared by the CGR, University of Liverpool, was sequenced on an Illumina MiSeq platform. This platform also uses sequencing by synthesis technology, as previously described. The read length was 50 bp, and approximately 80% of the total sequenced reads from each library mapped uniquely to the 4/74 reference genome. There were between approximately 3 million and 5 million uniquely mapped reads in each library.

Table 2.3 Strains and Growth Conditions used for HiSeq RNA-seq

Strain	<b>Growth Condition</b>	Description of Growth	Replicates <sup>a</sup>	Sequencing Run
wild-type	mid exponential phase (MEP)	growth to OD <sub>600</sub> 0.3 in Lennox broth	3 technical of 1 biological	1
$\Delta dam$	mid exponential phase (MEP)	growth to $OD_{600}0.3$ in Lennox broth	1 biological	1
wild-type	early stationary phase (ESP)	growth to $OD_{600}2.0$ in Lennox broth	4 biological and 2 technical of 1 biological	1, 2, 3
Δfur	early stationary phase (ESP)	growth to 0D <sub>600</sub> 2.0 in Lennox broth	1 biological	1
$\Delta hilD$	early stationary phase (ESP)	growth to OD <sub>600</sub> 2.0 in Lennox broth	1 biological	1
$\Delta barA/sirA$	early stationary phase (ESP)	growth to $OD_{600}2.0$ in Lennox broth	1 biological	2
$\Delta fliZ$	early stationary phase (ESP)	growth to $OD_{600}2.0$ in Lennox broth	1 biological	2
$\Delta rpoN$	early stationary phase (ESP)	growth to 0D <sub>600</sub> 2.0 in Lennox broth	1 biological	3
$\Delta pnp$	early stationary phase (ESP)	growth to $OD_{600}2.0$ in Lennox broth	1 biological	2
$\Delta hfq$	early stationary phase (ESP)	growth to OD <sub>600</sub> 2.0 in Lennox broth	1 biological	2
wild-type	late stationary phase (LSP)	growth to OD6002.0 in Lennox broth + 6 hours further growth	3 biological	1, 2

Table 2.3 Strains and Growth Conditions used for HiSeq RNA-seq

$\Delta rpoE$	late stationary phase (LSP)	growth to OD6002.0 in Lennox broth + 6 hours further growth	1 biological	1
$\Delta rpoS$	late stationary phase (LSP)	growth to OD6002.0 in Lennox broth + 6 hours further growth	1 biological	2
wild-type	SPI2 inducing (inSPI2)	growth to $OD_{600}0.3$ in $PCN^{b}$ (pH 5.8, 1 mM $MgCl^{2+},0.4$ mM $P_{i})$	2 biological	1, 2
$\Delta phoB/R$	SPI2 inducing (inSPI2)	growth to $OD_{600}0.3$ in PCN (pH 5.8, 1 mM $MgCl^{2+}\!,0.4$ mM $P_i)$	1 biological	1
$\Delta slyA$	SPI2 inducing (inSPI2)	growth to $OD_{600}0.3$ in PCN (pH 5.8, 1 mM MgCl $^{2+}\!,0.4$ mM $P_i)$	1 biological	1
$\Delta ompR/envZ$	SPI2 inducing (inSPI2)	growth to $OD_{600}0.3$ in PCN (pH 5.8, 1 mM MgCl $^{2+}\!,0.4$ mM $P_i)$	1 biological	2
$\Delta phoP/Q$	SPI2 inducing (inSPI2)	growth to $OD_{600}0.3$ in PCN (pH 5.8, 1 mM MgCl $^{2+}\!,0.4$ mM $P_i)$	1 biological	2
$\Delta ssrA$	SPI2 inducing (inSPI2)	growth to $OD_{600}0.3$ in PCN (pH 5.8, 1 mM $MgCl^{2+}\!,0.4$ mM $P_i)$	1 biological	2
$\Delta ssrB$	SPI2 inducing (inSPI2)	growth to $OD_{600}0.3$ in PCN (pH 5.8, 1 mM MgCl $^{2+},0.4$ mM $P_i)$	1 biological	2
$\Delta ssrA/B$	SPI2 inducing (inSPI2)	growth to $OD_{600}0.3$ in PCN (pH 5.8, 1 mM $MgCl^{2+}\!,0.4$ mM $P_i)$	1 biological	2
$\Delta hilD$	SPI2 inducing (inSPI2)	growth to $OD_{600}0.3$ in PCN (pH 5.8, 1 mM $MgCl^{2+},0.4$ mM $P_i)$	1 biological	2
wild-type	micro-aerobic (MMA)	static growth to $OD_{600}0.3$ in MMA <sup>c</sup> (0.4% glycerol, 40 mM $C_4H_2Na_2O_4$ , 20 mM TMAO)	1 biological	3
$\Delta fnr$	micro-aerobic (MMA)	static growth to $OD_{600}0.3$ in MMA (0.4% glycerol, 40 mM $C_4H_2Na_2O_4$ , 20 mM TMAO)	1 biological	3

<sup>&</sup>lt;sup>a</sup> Replicates were performed for this study and by others <sup>b</sup> (Lober *et al.*, 2006), (Xu & Hensel, 2010)

<sup>&</sup>lt;sup>c</sup> (Miller, 1972)

#### 2.6.9 Mapping of RNA-seq data and differential expression analysis

Reads obtained from RNA-seq experiments were mapped to the 4/74 reference genome by Dr. Karsten Hokamp and Sathesh Sivasankaran (Trinity College Dublin) using the Segemehl mapping software (Hoffmann et al., 2009). Reads that do not map in a single chromosomal location (uniquely mapped reads) were truncated from the 3' end in stepwise manner by removing one nucleotide at a time until the read is mapped uniquely or until the read length reaches 20 nucleotides. Remaining reads were discarded (Sivasankaran, 2014). Data were normalised using the transcripts per million (TPM) method (Wagner et al., 2012). This method of measuring transcript abundance from high throughput sequencing data is closely-related to the widely used reads per kilobase per million method (RPKM) (Mortazavi et al., 2008), but removes the bias, of normalisation to the total number of reads mapped in each sequencing run, that the RPKM method introduces. Instead the TPM method measures transcript abundance by calculating the number of transcripts of a particular gene by dividing the number of nucleotides mapped to that gene by the length of the gene (tg). All of these numbers are summed to get the total number of transcripts represented by all the mapped reads (T). The transcript abundance for each gene is then calculated as number of transcripts per million transcripts according to the following formula:

$$TPM = \frac{t_g}{T} 10^6$$

A TPM value of 2 is the suggested cut-off value for determining if a gene is expressed or not (Wagner *et al.*, 2013), however a more conservative cut-off of TPM = 10 was chosen for this study based on TPM values of indicator genes which were previously shown not to be expressed under a particular condition, as previously described (Kröger et al., 2013). Differential expression of genes between WT and isogenic mutants was calculated from TPM values.

## 2.6.10 Analysis of RNA stability

Overnight cultures were diluted in 25 mL of pre-warmed media in a 250 mL flask with appropriate antibiotics and grown to the desired  $OD_{600}$ , as previously described. Four

 $OD_{600}$  units were harvested for RNA isolation, as previously described (Time 0). Rifampicin was added to the culture at a final concentration of 150 µg/mL. Four  $OD_{600}$  units of culture were harvested for RNA isolation at regular intervals following the addition of Rifampicin. Rifampicin interacts with bacterial RNAP and prevents transcript elongation, thus addition of Rifampicin prevents *de novo* RNA synthesis. Cellular concentrations of an RNA transcript are dependent on the rate of transcription and the rate of transcript degradation. Prevention of *de novo* RNA synthesis, using Rifampicin, allows us to determine the rate of transcript degradation. Cellular transcript concentrations were measured using northern blotting of RNA isolated at each timepoint. Densitometry analysis was subsequently used to quantify the northern blot signal. Signal intensity was plotted against time (in minutes, after addition of Rifampicin) to determine transcript half-lives. Densitometry analysis was performed using ImageJ software (Schneider *et al.*, 2012).

# 2.7 Analysis of proteins

# 2.7.1 Preparation of whole cell lysates for analysis of cellular proteins

Overnight cultures were diluted in 25 mL of pre-warmed media in a 250 mL flask with appropriate antibiotics and grown to the desired  $OD_{600}$  as previously described. 0.2  $OD_{600}$  units were harvested by centrifugation at 4°C for 5 minutes at 13,000 rpm. Cell pellets were washed once in 1 mL of PBS. Cell pellets were finally re-suspended in 25  $\mu$ L of nuclease-free water and mixed with 5X Laemmli buffer (0.25 M Tris-HCl pH 6.8, 15% SDS, 50% glycerol, 25%  $\beta$ -mercaptoethanol and bromophenol blue) to a final concentration of 1X Laemmli buffer (Laemmli, 1970). Samples were boiled for 10 minutes at 99°C and stored at -20°C.

#### 2.7.2 Preparation of culture supernatants for analysis of secreted proteins

Overnight cultures were diluted in 25 mL of pre-warmed media in a 250 mL flask with appropriate antibiotics and grown to the desired  $OD_{600}$  as previously described. Ten OD units were harvested by centrifugation at 4°C for 40 minutes at 4,000 rpm. Supernatants were filtered through a 0.45  $\mu$ M pore size filter into a clean tube. Ice-cold trichloroacetic

acid was added to a final concentration of 10% to precipitate proteins from the culture supernatants and samples were centrifuged at  $4^{\circ}$ C for 40 minutes at 4,000 rpm. Supernatants were removed and protein pellets were re-suspended in 1 mL of ice-cold acetone, moved to a 1.5 mL tube and centrifuged at  $4^{\circ}$ C for 10 minutes at 8,000 rpm. Pellets were finally re-suspended in 30  $\mu$ L of nuclease-free water and mixed with 5X Laemmli buffer to a final concentration of 1X Laemmli buffer. Samples were boiled for 5 minutes at 99°C and stored at -20°C.

#### 2.7.3 SDS polyacrylamide gel electrophoresis (SDS-PAGE)

SDS polyacrylamide gels were used for the separation of denatured proteins based on their molecular weight using a discontinuous gel system (Sambrook, 1989). Resolving gels were made from a stock solution of 40% acrylamide/bis-acrylamide (37:5:1) to between 10% and 15% final concentration and contained 0.375 M Tris-Cl (pH 8.8) and 0.1% SDS. Gels were polymerised using 0.1% (w/v) ammonium persulfate (APS) and 0.01% (v/v) of tetramethylethylenediamine (TEMED). The separating gel was overlaid with 1 mL of isopropanol to aid polymerisation. After polymerisation the isopropanol layer was poured off and excess isopropanol was removed with filter paper. Stacking gels contained a final concentration of 5% acrylamide, 0.082 M Tris-Cl (pH 6.8) and 0.1% SDS. Stacking gels were also polymerised with 0.1% (w/v) APS and 0.01% (v/v) TEMED. Unless otherwise stated 0.1 OD<sub>600</sub> unit of whole cell lysates or 5 OD<sub>600</sub> units of culture supernatants were loaded on polyacrylamide gels. EZ Run prestained Rec protein ladder was used as a protein molecular weight marker. Gels were run at a constant voltage of 90 V in 1X running buffer (25 mM tris base, 0.19 M glycine, 0.1% SDS) until the dye front reached the bottom of the gel. Electrophoresed proteins were then analysed by Coomassie staining or western immunoblotting.

#### 2.7.4 Coomassie staining of polyacrylamide gels

After gel electrophoresis, total protein profiles could be analysed using Coomassie brilliant blue R-250 stain (3 mM Coomassie R-250, 45% methanol, 10% acetic acid). Gels were rocked in Coomassie staining solution for at least 3 hours at room temperature. Gels were de-stained by rocking in Coomassie de-staining solution (45% methanol, 10%).

acetic acid) at least twice until the background had become clear. Protein profiles were visualised with white light using an ImageQuant LAS 4000 Imager.

#### 2.7.5 Western immunoblotting

Western immunoblotting was used to detect specific proteins. Gels containing electrophoretically separated polypeptides, nitrocellulose membrane (0.45  $\mu$ M pore size) and filter paper were immersed in 1X transfer buffer (25 mM tris, 0.19 M glycine). Proteins were transferred from the gel to nitrocellulose membrane using a semi-dry blotting apparatus at a constant amplitude of 300 mA for 90 minutes at 4°C.

To avoid non-specific protein binding, membranes were blocked for 1 hour at room temperature with rocking in blocking buffer (5% skimmed milk powder, PBS, 0.01% Tween-20). Primary antibodies were diluted as shown in Table 2.4 in blocking buffer and incubated with the membrane overnight with rocking at 4°C. Primary antibody solutions were stored at -20°C for re-use. DnaK served as a loading control and a fresh dilution of anti-DnaK antibody was prepared for each western blot and mixed with the experimental antibody. Membranes were washed 4 times for a total of 40 minutes in cold PBST washing buffer (PBS, 0.01% Tween-20). Antigen-primary antibody complexes were detected using horseradish peroxidase (HRP) conjugated secondary antibody, which was also diluted in blocking buffer and incubated with the membrane for 90 minutes with rocking at room temperature. The membrane was washed three times for a total of 30 minutes in cold PBST, followed by a final wash for 15 minutes in PBS. Chemiluminescent signal detection was carried out using the ECL chemiluminescent HRP substrate (Pierce) containing a 1:1 mixture of luminol and peroxide reagents resulting in light emission, when cleaved by horse radish peroxidase. The light emission was detected using an ImageQuant LAS 4000 imager.

Table 2.4 Antibodies used for western immunoblot analysis

Antibody	Source	Dilution	Supplier
Monoclonal anti-FLAG M2	Mouse	1:10,000	Sigma-Aldrich (F3165)
Monoclonal anti-GFP	Mouse	1:1,000	Roche (11814460001)
Monoclonal anti-DnaK (E. coli)	Mouse	1:10,000	Enzo Life Sciences (ADI-SPA-880)
Monoclonal anti-RpoS	Mouse	1:10,000	Neoclone (W0009)
anti-Mouse IgG (H+L) HRP conjugate	Goat	1:3,000	Biorad (172-1011)

### 2.8 Chromatin Immunoprecipitation (ChIP)

#### 2.8.1 Preparation of cross-linked lysates

Chromatin immunoprecipitation (ChIP) was carried out as previously described (Dillon et al., 2010) with some alterations. Overnight cultures were diluted in 25mL of pre-warmed media in a 250 mL flask with appropriate antibiotics and grown to the desired OD<sub>600</sub>. Between 6 and 12 OD units were harvested by centrifugation at room temperature at 4,000 rpm for 8 minutes. The cells were washed in 50 mL of pre-warmed (37°C) PBS. Protein-DNA complexes were cross-linked by adding formaldehyde to a final concentration of 1% in a drop-wise manner with gentle stirring at room temperature for 30 minutes. Cross-linking reactions were quenched by the addition of ice-cold glycine to a final concentration of 0.125 M for 5 minutes with gentle stirring at room temperature. The cross-linked cells were harvested by centrifugation at 4°C at 4,000 rpm for 8 minutes and were re-suspended in 600 µL of lysis buffer (50 mM Tris-HCl pH 8.1, 10 mM EDTA, 1% SDS, 1X protease inhibitor tablet stock) and incubated on ice for 10 minutes. 1.4 mL dilution buffer (20 mM Tris-HCl pH 8.1, 150 mM NaCl, 2 mM EDTA, 1% Trition X-100, 0.01% SDS, 1X protease inhibitor tablet stock) was added and the chromatin was sonicated on ice to reduce the average DNA fragment length to approximately 500 bp using an MSE Soniprep sonicator (Sanyo). Sonication generally proceeded as 9 bursts at an amplitude of 10 µM for 30 seconds/burst, followed by 1 minute of cooling on ice after each burst. 5 µL of sonicated chromatin was run on an agarose gel to check sonication efficiency. Cellular debris was pelleted by centrifugation at 13,000 rpm for 10 minutes at 4°C. The supernatant was moved to a 15 mL tube and 1 mL of dilution buffer was added. Sheared chromatin was then stored at -80°C.

#### 2.8.2 Immunoprecipitation

To remove non-specific antibody binding sites, the chromatin was first pre-cleared by adding 50  $\mu g$  of non-species specific IgG, typically normal rabbit IgG (Millipore). The chromatin was incubated, in 15 mL tubes, for 1 hour at 4°C on a rotating wheel and 100  $\mu L$  of protein G-agarose bead suspension was added. The chromatin was incubated for a further 3 hours at 4°C with rotation. Beads were pelleted at 4,000 rpm for 4 minutes at 4°C. 200  $\mu L$  of pre-cleared chromatin was removed and stored at -20°C as "Input" DNA for downstream analysis. The remainder of the pre-cleared chromatin was used to set up Immunoprecipitation (IP) reactions in 2 mL tubes. A "mock" IP reaction containing 1350  $\mu L$  of chromatin and 10  $\mu g$  of species specific IgG antibody, typically normal mouse IgG (Millipore) was set up to measure background levels of DNA binding to antibodies and beads. A typical experimental IP reaction contained 1350  $\mu L$  of chromatin and 10  $\mu g$  of monoclonal mouse anti-FLAG M2 antibody (Sigma). IP reactions were incubated overnight at 4°C on a rotating wheel. 50  $\mu L$  of protein G-agarose bead suspension was added to each IP sample and incubation continued for 3 hours at 4°C on a rotating wheel.

#### 2.8.3 Washing the Protein G-agarose beads and elution of DNA

The beads containing the bound antibody-protein-DNA complexes were carefully washed as follows: beads were pelleted by centrifugation at 7400 rpm for 2 minutes at  $4^{\circ}$ C and incubation on ice for 1 minute. The supernatant was removed by pipetting and 750  $\mu$ L of each wash buffer was added, followed by vortexing of the samples and pelleting of the beads. The beads were washed twice in IP wash buffer 1 (20 mM Tris-HCl pH 8.1, 50 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% SDS), followed by one wash in IP wash buffer 2 (10 mM Tris-HCl pH 8.1, 250 mM LiCl, 1 mM EDTA, 1% NP-40, 1% deoxycholic acid) and 2 washes in TE buffer pH 8.0. Antibody-protein-DNA complexes were eluted from the beads at room temperature by adding 225  $\mu$ L of elution buffer (100 mM NaHCO<sub>3</sub>, 1% SDS) followed by vortexing and pelleting of the beads twice. Both eluates were combined in the same tube.

#### 2.8.4 Reversal of cross-links

Input and IP samples were treated with 5 ng/ $\mu$ L RNase A (Sigma) and 0.3 M NaCl and incubated at 65°C for at least 6 hours or overnight. Protein-DNA cross-links were disrupted by treating Input and IP samples with 9  $\mu$ g of Proteinase K at 45°C for at least 3 hours or overnight.

#### 2.8.5 DNA extraction and analysis

DNA was extracted from Input and IP samples by standard phenol-chloroform extraction followed by ethanol precipitation with yeast tRNA and glycogen as co-precipitants. Briefly, 10  $\mu$ g of yeast tRNA was added to each sample prior to addition of a 1:1 mixture of phenol (pH 8.0) and chloroform. Samples were vortexed and centrifuged at 13,000 rpm for 5 minutes at room temperature. The aqueous phase was moved to a new tube and 500  $\mu$ L of chloroform was added to each sample. The samples were vortexed and centrifuged as before and the aqueous phase was moved to a new tube. DNA was precipitated by the addition of 5  $\mu$ g of glycogen, 5  $\mu$ g of yeast tRNA, sodium acetate (pH 5.2) and 100% ethanol. Samples were incubated at -80°C for 1 hour. The DNA was pelleted by centrifugation at 13,000 rpm for 20 minutes at 4°C followed by rinsing in ice-cold 70% ethanol and centrifugation at 13,000 rpm for 15 minutes. DNA pellets were allowed to air-dry for 15 minutes. Input DNA was re-suspended in 100  $\mu$ L nuclease free water and IP DNA was re-suspended in 50  $\mu$ L nuclease free water at 37°C with shaking at 900 rpm for 1 hour.

Input and IP DNA was further diluted 50-fold and 5-fold respectively for analysis by quantitative PCR as previously described. DNA was quantified relative to a standard curve of chromosomal DNA. The quantity of immunoprecipitated DNA is relative to specific protein binding in that region and was calculated as a fraction of the starting amount of DNA (Input). The mock immunoprecipitate was subtracted from the experimental immunoprecipitate and compared to a control region which was negative for specific transcription factor binding according to the formula:

Experimental IP of promoter X	Mock IP of promoter X				
Experimental Input of promoter X	 Mock Input of promoter X	•			

### 2.9 GFP-based two-plasmid validation assay for sRNA targets

#### 2.9.1 Plasmids

The superfolder GFP-based two-plasmid assay (Urban & Vogel, 2007, Corcoran et al., 2012) is a reporter system, used for the validation of predicted mRNA targets of sRNAs. The two-plasmid assay involves the use of fusion plasmids to demonstrate direct interaction between sRNAs and their putative interaction sites on target mRNAs. Fusion plasmids, pXG10-(sf) and pP<sub>L</sub>, detailed in Table 2.2, were used in this study. The 5' UTR and start of the coding sequence (CDS) of a putative target mRNA was cloned in-frame with the superfolder GFP sequence, under control of the constitutive P<sub>LtetO</sub> promoter on the pXG10 plasmid. The full-length sRNA was cloned onto the pP<sub>L</sub> plasmid, under the control of the constitutive P<sub>LlacO</sub> promoter. The pP<sub>L</sub> plasmid, JV300, expressing a nonsense RNA, derived from the *rrnB* terminator, from the P<sub>LlacO</sub> promoter was used as a control plasmid for comparison of GFP expression (Sittka et al., 2007).

#### 2.9.2 The two-plasmid target validation assay

*E. coli* TOP10 cells were chemically transformed (section 2.4.2) with the pXG10-(sf) fusion and the pP<sub>L</sub> fusion, containing the sRNA of interest or the nonsense RNA. Overnight cultures of the resultant double transformants were grown, using selective antibiotics for each plasmid (Table 2.2). GFP signal output is a measure of the interaction between the sRNA and mRNA fusion, reporting the sRNA-mediated target regulation. GFP fluorescence can be measured by growth on agar plates, flow cytometry or western blot (Urban & Vogel, 2007). Western blotting (section 2.7.5) was used for measurement of GFP fluorescence in this study. Higher GFP fluorescence, following co-expression of the pXG10-(sf) fusion strain with the control pP<sub>L</sub>, compared to the sRNA-expression pP<sub>L</sub> vector, indicates that the sRNA represses the target mRNA. Lower GFP fluorescence following co-expression of the pXG10-(sf) fusion with the control pP<sub>L</sub> vector, compared to the sRNA-expressing pP<sub>L</sub> vector, indicates that the sRNA activates expression of the target mRNA.

### 2.10 Bioinformatic analysis

#### 2.10.1 Calculation of Pearson correlation coefficient

Correlative analysis was performed to compare gene expression across datasets, such as between RNA-seq runs, in wild-type and mutant strains grown under 5 environmental conditions (this study), in wild-type grown under 22 environmental conditions (Kröger et al., 2013). The Pearson correlation coefficient was calculated using Excel software (Microsoft Office). Typically, two datasets with a Pearson correlation coefficient >0.7 were considered to have a strong positive linear relationship.

#### 2.10.2 Generation of a position-specific scoring matrix (PSSM)

PSSMs or position weight matrices are used to define specific DNA binding sequences of transcription factors and can be used to predict new binding sites of that TF (Stormo & Zhao, 2010). PSSMs were generated from characterised TF binding sites (as discussed in section 3.2.8), by assigning a score to each possible base at each position within the binding site. The combination of the scores of each base that correspond to a binding site give an overall score for that binding site (Stormo & Zhao, 2010). The PSSM was normalised to the average G/C content in the S. Typhimurium chromosome. The PSSM was used to scan regions or promoters of interest to predict new TF binding sites using pattern searching software available from <a href="www.rsat.ulb.ac.be">www.rsat.ulb.ac.be</a>. Typically the lowest scoring TF binding site from the characterised TF binding sites was used as a threshold score for prediction of new TF binding sites. Predicted binding sites with scores lower than the threshold were considered to be false positives.

#### 2.10.3 List of online tools and programmes used for bioinformatic analysis in this study

- Integrated Genome Browser (IGB) Visualisation of RNA-seq reads and other genome features (Nicol et al., 2009).
- JBrowse Visualisation of RNA-seq reads and other genome features (Skinner et al., 2009).

- GeneSpring 7.3 Gene expression data analysis and generation of cluster heatmaps (Agilent Technologies).
- SalCom S. Typhimurium mRNA and sRNA expression profiles based on RNA-seq data from 20 environmental conditions and within murine macrophages (Kröger et al., 2013, Srikumar et al., 2014).
- Nupack Prediction of RNA secondary structures (Zadeh *et al.*, 2011).
- TargetRNA2 Prediction of targets for bacterial sRNAs (Kery et al., 2014).
- IntaRNA Prediction of RNA-RNA interactions (Busch *et al.*, 2008).
- CopraRNA Whole genome prediction of sRNA targets (Wright et al., 2013).
- BLAST Local sequence alignment and identification of regions of sequence similarity (NCBI).
- Multalin Generation of multiple sequence alignments (Corpet, 1988).
- Cytoscape Generation of transcriptional regulatory network (Shannon et al., 2003).
- Regulatory Sequence Analysis Tools (RSAT) Scanning of PSSMs to identify regulatory motifs (Thomas-Chollier *et al.*, 2011).
- WebLogo Generation of stacked sequence logos based on multiple sequence alignment (Crooks *et al.*, 2004).
- ImageJ Densitometry analysis of northern and western blot signals (Schneider et al., 2012).
- Venny Generation of Venn diagrams.

# **Chapter 3**

# Global analysis of regulatory inputs on Salmonella Typhimurium small RNAs

#### 3.1 Introduction

#### 3.1.1 Overview of the study

S. Typhimurium co-ordinates a sophisticated and complex transcriptional network, controlled by regulatory proteins integrating combinations of regulatory inputs in order to establish successful infection (Figure 1.3) (Yoon et al., 2011, Osborne et al., 2009). In recent years the discovery of large numbers of small non-coding RNAs in bacteria has challenged our traditional views on gene regulation and regulatory networks. It has been suggested that every gene in E. coli is directly or indirectly regulated by an sRNA (Holmqvist et al., 2012). The wide scope of functions that have already been attributed to sRNAs highlights the importance of sRNA-mediated post-transcriptional gene regulation. In bacteria, sRNAs have been shown to be involved in various processes including homeostasis of the outer membrane (Papenfort et al., 2006, Pfeiffer et al., 2007, Frohlich et al., 2012), amino acid metabolism (Sharma et al., 2011), anaerobic growth (Boysen et al., 2010, Durand & Storz, 2010), iron homeostasis (Masse & Gottesman, 2002), toxin/antitoxin systems (Vogel et al., 2004) and virulence (Padalon-Brauch et al., 2008, Gong et al., 2011, Santiviago et al., 2009). The sRNA-mediated regulation of expression of TF genes indicates that sRNAs are involved at multiple levels in the control of gene expression (Gopel & Gorke, 2012, Mandin & Guillier, 2013). The wide-ranging functions of sRNAs require that expression of sRNAs must be carefully regulated in order to maximise sRNA function, while avoiding incurring fitness costs as a result of inappropriate gene expression. It is clear that sRNAs and regulatory proteins interact and form densely interconnected regulatory networks (Gopel & Gorke, 2012, Mandin & Guillier, 2013), but to our knowledge, there have been no large-scale investigations to identify the protein factors involved in controlling sRNA expression.

The aim of this study is to identify the virulence-associated regulatory inputs that control expression of sRNAs in *S*. Typhimurium, with a view to understanding the role played by sRNAs in *S*. Typhimurium pathogenesis.

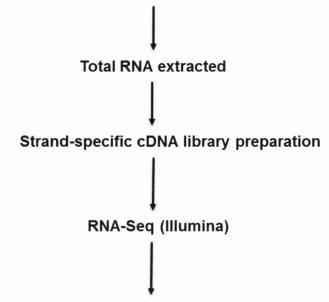
This work has mainly focussed on transcriptional regulation of sRNAs by using RNA-seq-based transcriptomic analysis of 17 important regulatory mutants under 5 environmental conditions to identify differential sRNA gene expression. A panel of virulence-associated transcription factors, two-component systems and sigma factors,

grown under infection-relevant conditions, were selected. A secondary aim of this study was to investigate the roles of the RNA chaperone Hfq and the exoribonuclease PNPase in post-transcriptional regulation of sRNA expression. The transcriptomes of all mutant strains were analysed by RNA-seq and compared to wild-type grown under the same conditions. A general workflow of the main experimental procedure is outlined in Figure 3.1.

The virulence-associated regulatory proteins (discussed in section 1.2) perform a wide variety of functions. Some of the regulatory proteins form part of the core ancestral S. Typhimurium chromosome, while others were horizontally acquired on pathogenicity islands. Figure 3.2 is a BLAST atlas of the S. Typhimurium SL1344 genome, showing the location and conservation of each of the genes encoding the regulatory proteins under investigation in this study. SL1344 is derived from the wild-type strain used in this study (4/74) and there are only 8 SNP differences between SL1344 and 4/74 (Kröger et al., 2012). The genomes of three S. Typhimurium strains (4/74, LT2 and D23580), S. Enteritidis, S. Typhi, S. bongori and E. coli K-12 were compared to the reference SL1344 genome by BLAST analysis. Increasing colour intensity at a locus indicates sequence conservation in the comparator genome to the reference genome at this locus, while absence of a colour indicates absence of that locus in the comparator genome. The SPI1 island, and the SPI1-encoded TF HilD, are present in all Salmonella genomes but are absent from E. coli K-12. The SPI2 island, and SPI2-encoded TCS SsrA/B, are present in all S. enterica genomes but are absent from S. bongori and E. coli. Fourteen sRNA genes are also indicated on the BLAST atlas (labelled in blue). The genomic location and conservation of these sRNAs will be discussed in greater detail in sections 3.2.10 and 3.2.11.

The regulators presented in Figure 3.2 represent a diverse range of proteins that play important roles for survival and establishment of intracellular infection by *S*. Typhimurium. Analysis of sRNA expression in mutant strains lacking this panel of regulatory proteins provides a clearer picture of the complex regulatory interactions that exist between the transcriptional and post-transcriptional elements of *S*. Typhimurium gene regulation, and may also provide clues about the function of uncharacterised sRNAs. This type of analysis is not possible through interrogation of individual components of these regulatory networks (Hebrard et al., 2011, Golubeva et al., 2012).

WT & mutant S. Typhimurium strains grown under relevant in vitro conditions



Mapping, TPM calculation and differential expression analysis

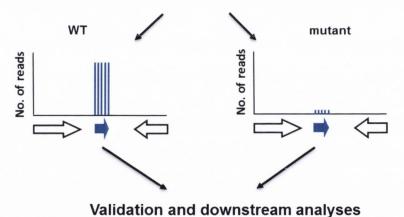


Figure 3.1 Workflow of general experimental procedures for investigating the regulatory inputs on sRNA genes.

Wild-type (WT) and mutant strains were grown under appropriate environmental conditions and total RNA was extracted by me. cDNA library preparation and RNA-seq were performed by Vertis Biotechnologie AG. Mapping of sequence reads to the reference ST4/74 genome and calculation of TPM values were performed by Dr. Karsten Hokamp and Dr. Sathesh Sivasankaran (Trinity College Dublin). Differential gene expression analysis, validation and downstream analyses were performed by me.

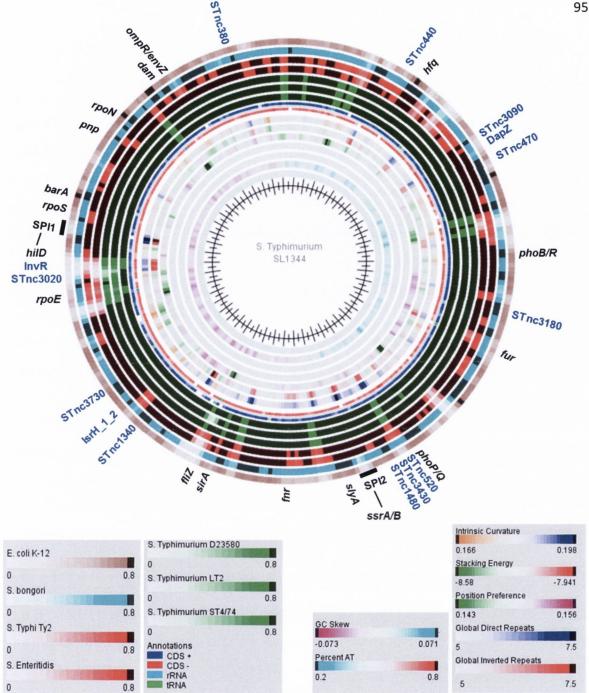


Figure 3.2 BLAST atlas of the S. Typhimurium SL1344 genome showing the location and conservation of genes important for this study

A BLAST atlas using S. Typhimurium SL1344 as a reference genome was previously generated (Kröger et al., 2012) using software available from www.cbs.dtu.dk. indicated in the figure key, the innermost circles represent DNA parameters, such as AT or GC content, intrinsic curvature, global direct repeats and global inverted repeats. The plus and minus coding strands are shown as blue and red circles, respectively. The results of the BLAST analysis of 7 bacterial genomes to the reference genome are shown in the outer circles: the green circles represent the S. Typhimurium 4/74, LT2 and D23580 genomes. The S. Enteritidis and S. Typhi genomes are shown in red. The S. bongori genome is shown in turquoise. The E. coli K-12 genome is shown in brown. Increasing intensity of colour indicates increased similarity of the locus to the reference genome. Absence of colour indicates the absence of the locus. Important regions or genes are indicated at their location outside the BLAST lanes. Protein-coding genes are labelled in black, sRNA genes are labelled in blue.

#### 3.1.2 RNA-seq

RNA-seq is the high-throughput sequencing of cDNA libraries. Currently, direct sequencing of RNA is not yet possible for bacterial transcriptomic analysis, although direct RNA sequencing has been used for profiling the transcriptome of the yeast Saccharomyces cerevisiae (Ozsolak et al., 2009). cDNA sequencing or RNA-seq is a method of identifying and quantifying which genes are being actively transcribed within the cell on a global scale. This technology provides a highly sensitive and dynamic method of transcriptome profiling at single nucleotide resolution (Ozsolak & Milos, 2011). Depletion methods may be used to avoid saturation of the data by pervasive RNA species, such as rRNA, while enrichment methods may be used to focus on a particular RNA species, for example size exclusion methods can be used to focus on small transcripts (Liu & Camilli, 2011). The cDNA libraries used for RNA-seq in this study were generated from total RNA by Vertis Biotechnologie AG, Freising, Germany. RNA-seq was performed using an Illumina Hiseq 2000 platform as previously described (section 2.6.7 and 2.6.8). No methods of depletion were used during cDNA library preparation in this study, but analysis was limited to the sequencing reads that mapped to a single location on the chromosome (uniquely mapped reads), and reads mapping to many rRNA genes or other paralogous genes were removed. A recent systematic analysis of the effects of sequencing depth on discovery of rare transcripts demonstrated that sequencing of 5-10 million reads of non-rRNA fragments provides adequate coverage of a bacterial transcriptome and is sufficient to detect most expressed genes (Haas et al., 2012). Therefore no enrichment for sRNAs was required in this study as the depth of sequencing achieved permitted the detection of even rare transcripts. Total DNase I digested RNA was fragmented and converted to cDNA fragments. Short bar-code sequences were ligated to the RNA fragments in pre-determined orientations to allow for the generation of strand specific sequencing data, which is particularly useful in identifying antisense-encoded sRNAs (Levin et al., 2010). The development of RNA-seq analysis pipelines (Sivasankaran, 2014), and the use of TPM values to quantify gene expression (as described in section 2.6.9) permitted information to be managed and distilled from these large datasets.

These methods provided reproducible and comprehensive gene expression datasets that have detected new sRNAs and identified thousands of transcriptional start sites (Kröger et al., 2012, Kröger et al., 2013). The use of RNA-seq to investigate the transcriptomes of

19 regulatory mutants is helping to expand our knowledge of the coding and non-coding portions of *Salmonella* virulence-associated regulons. Validation of this relatively novel method of mutant transcriptome profiling, as well as the key findings and highlights of the global analysis of regulatory inputs to *Salmonella* genes, especially sRNA genes, are presented below. Absolute expression values (in TPM) for every chromosomal *S*. Typhimurium 4/74 gene in every mutant and wild-type comparator strain, sequenced in this study, are available in Appendix I.

#### 3.2 Results

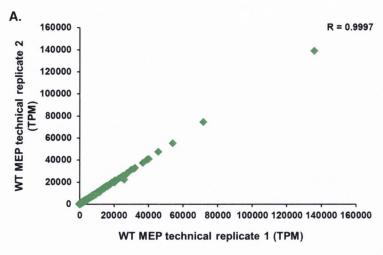
#### 3.2.1 Reproducibility of RNA-seq data

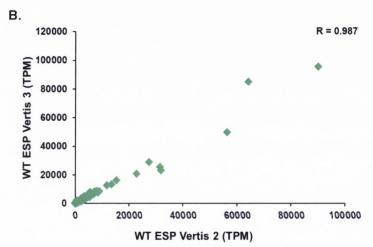
Inherent biological variability and random variation, such as that introduced during sample preparation, are features of any assay used in microbiology. Replication of experiments is necessary to identify and minimise the variation in experimental systems (Quackenbush, 2002), and to provide statistical significance for the acquired data. Additionally, the use of an RNA-seq-based approach for transcriptomics is a relatively recent development, thus it is necessary to prove that this technology can be trusted to accurately and reproducibly detect differential gene expression. However, using RNA-seq to determine the transcriptome of 19 regulatory mutants and their wild-type comparators in duplicate or triplicate would be a prohibitively expensive experiment.

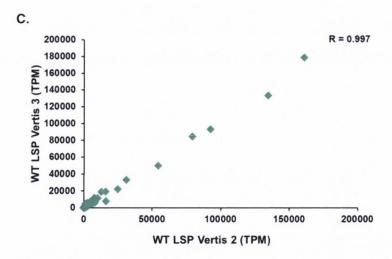
The sequencing data used in this study were acquired in 3 separate sequencing runs using independent RNA samples. It was, therefore, important to quantify the reproducibility of the results between sequencing runs in order to compare gene expression across all strains. Independent biological replicates of wild-type (WT) cells grown to ESP, LSP and inSPI2 were each included in the first and second sequencing runs (Vertis 2 and Vertis 3 respectively), while technical replicates of the wild-type MEP sample were sequenced in the first sequencing run. The reproducibility and robust nature of the RNA-seq-based transcriptomic data set were confirmed using correlative analysis of the technical replicates from samples sequenced in the same sequencing run but in different Illumina flow cells, and between independent biological replicates sequenced in different sequencing runs, as previously described (Kröger et al., 2013) (section 2.10.1).

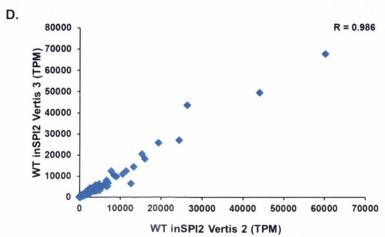
The gene expression values between the technical replicates of WT MEP (Figure 3.3 A) show a high level of correlation (Pearson correlation coefficient: 0.9997). There is also a high level of correlation between biological replicates of WT ESP, WT LSP and WT inSPI2 (Figure 3.3 B, C and D; correlation coefficients: 0.987, 0.997 and 0.986 respectively). The high degree of correlation between technical and biological replicates of wild-type samples from 4 environmental conditions demonstrates the reproducibility of the RNA extraction, cDNA library preparation and RNA-seq methods. During downstream analyses, independently extracted RNA was used to validate many RNA-seq-based findings by northern blot and quantitative PCR.

Figure 3.3 High correlations of RNA-seq data from technical and independent biological replicates









# Figure 3.3 High correlations of RNA-seq data from technical and independent biological replicates

**A.** RNA-seq-based transcriptomic data from wild-type cells grown to MEP show a high level of correlation (close to 1) with a technical replicate sequenced in the same sequencing run but in a different Illumina flow cell. **B, C, D.** RNA-seq-based transcriptomic data from wild-type cells, grown to ESP, LSP and inSPI2 respectively, show a high level of correlation (close to 1) with independent biological replicates from different sequencing runs.

#### 3.2.2 Construction and sequencing of chromosomal deletion mutants

As discussed in chapter 1, the regulators chosen for mutagenesis and transcriptomic analysis in this study play an important role in S. Typhimurium virulence and/or survival during S. Typhimurium infection. All mutant strains were generated using the Datsenko and Wanner method of gene deletion as described in section 2.5.1 (Datsenko & Wanner, 2000). This method involves replacing the gene of interest, or part of the gene of interest, with an antibiotic resistance cassette by homologous recombination. The recombination event is mediated by the plasmid-encoded recombinase system of Bacteriophage  $\lambda$  (Murphy, 1998). The resistance cassette is amplified from a plasmid using oligonucleotides which also contain sequences homologous to the DNA on either side of the gene to be deleted. FRT sites flank the plasmid-encoded resistance gene to allow for removal of the antibiotic resistance gene from the resultant mutant strain using the pCP20 plasmid (Cherepanov & Wackernagel, 1995). Removal of the antibiotic resistance gene leaves an 82-85 nucleotide scar which contains an idealised ribosome binding site and a start codon to allow for expression from the downstream gene.

Using previously published transcriptomic and transcriptional start site (TSS) data from wild-type ST4/74 grown under 22 environmental conditions (Kröger et al., 2013), it was determined which downstream genes might be affected in the panel of deletion mutants used in this study. To avoid polar effects, resistance cassettes were removed from mutant strains if downstream genes were transcribed from the TSS of the deleted gene. Typically, entire coding regions were removed but the exceptions are detailed below and in Appendix IX. Figure 3.4 shows RNA-seq data from the regions surrounding each deleted gene in the corresponding mutant strain and wild-type comparator grown under the same conditions (visualised using the Integrated Genome Browser IGB) (Nicol et al.,

2009). It is evident from Figure 3.4 that each gene to be deleted was expressed in the wild-type strain under the conditions used in this study, and that, in most cases, construction of this panel of regulatory mutants did not cause significant polar downstream effects. Cases in which expression of the downstream gene is affected in the mutant strain are discussed below.

hilA, the gene immediately downstream of the hilD gene, is not expressed in the  $\Delta hilD$  mutant strain (Figure 3.4 B). However, this effect is regulatory rather than a technical effect of the deletion of the hilD gene as the hilA gene is transcriptionally activated by HilD (Ellermeier & Slauch, 2007). Additionally, the resistance cassette was removed from the  $\Delta hilD$  mutant strain and hilA is transcribed from its own TSS (Kröger et al., 2013).

The 3' end of the sirA coding sequence overlaps with the TSS of the downstream uvrC gene (Figure 3.4 D). To maintain an intact uvrC TSS, only 432 bp of the sirA gene were removed and the 3' of the sirA gene was not deleted. However, expression of the uvrC gene is reduced approximately 4-fold in the  $\Delta barA/sirA$ -432 mutant. The reduction in expression of uvrC may be a result of direct or indirect regulation by the BarA/SirA two-component system. However in E. coli, uvrY and uvrC are expressed as an operon and the majority of uvrC expression is driven from the promoter upstream of uvrY (Stark & Moses, 1989) and this is also likely to be the case for the Salmonella uvrYC genes. The *uvrC* gene encodes an endonuclease that is part of the UvrABC system which detects and aids in the repair of DNA damage by nucleotide excision. While there is a constitutive basal level of uvrC expression in E. coli, the gene is mostly repressed by the LexA repressor and becomes activated as part of the SOS response (Stark & Moses, 1989). The  $\Delta barA/sirA$ -432 mutant and wild-type comparator were grown in rich media to early stationary phase without mutagens that cause DNA damage or induce the SOS response, so high levels of UvrC expression may not be necessary under this condition. Differentially expressed genes in the  $\Delta barA/sirA$ -432 mutant compared to wild-type can therefore be considered to be directly or indirectly regulated by the BarA/SirA TCS, rather than being due to damaged and unrepaired DNA as a result of the loss in uvrC expression.

The 5' and 3' ends of the *pnp* gene were left intact to avoid disrupting expression of the downstream *nlpI* gene (Figure 3.4 H). A section of 1.95 kb of the *pnp* gene was removed.

Expression of *nlpI* increases approximately 4-fold in absence of PNPase. *pnp* and *nlpI* form a polycistronic transcript, and the effect of increased *nlpI* expression in various non-polar *pnp* mutants has been reported (Rouf et al., 2011). The *pnp-nlpI* transcript is negatively auto-regulated via PNPase-mediated degradation (Carzaniga et al., 2009), resulting in increased levels of *nlpI* transcripts in the absence of PNPase expression. Thus the increase in *nlpI* expression is likely to be a regulatory effect, rather than the effect of polarity.

A 335 bp region of the slyA gene was removed to generate the  $\Delta slyA$  mutant strain used in this study (Figure 3.4 L). The 5' end of the gene was left intact to avoid disrupting expression of the downstream anmK gene which is transcribed from the same promoter. Expression of anmK is not significantly affected in the  $\Delta slyA$  mutant strain. The 3' end of the slyA gene was left intact to avoid disrupting the slyB gene which is encoded on the opposite strand and overlaps slyA at the 3' end of both transcripts. slyB expression is reduced approximately 4-fold in the  $\Delta slyA$  strain, however this is likely to be a regulatory effect rather than caused by polarity. SlyA does not regulate slyB expression directly (Perez et al., 2008) but SlyA activates phoP (Song et al., 2008), which in turn activates slyB (Lejona et al., 2003) and this cascade may explain the reduction in slyB expression in the  $\Delta slyA$  mutant.

Expression of *ssrA* is reduced approximately 4-fold in the Δ*ssrB* mutant (Figure 3.4 O). This reduction in *ssrA* expression is also likely to be a regulatory effect rather than a technical effect, as SsrB binds and activates expression from the *ssrA* promoter (Tomljenovic-Berube et al., 2010). Expression of the gene downstream of *ssrAB* locus, *orf242*, is also reduced approximately 4-fold in the absence of SsrA and SsrB (Figure 3.4 P). *orf242* is transcribed from its own promoter and the reduction in *orf242* expression is likely to be regulatory as SsrB binds in the intergenic region upstream of *orf242* (Tomljenovic-Berube et al., 2010).

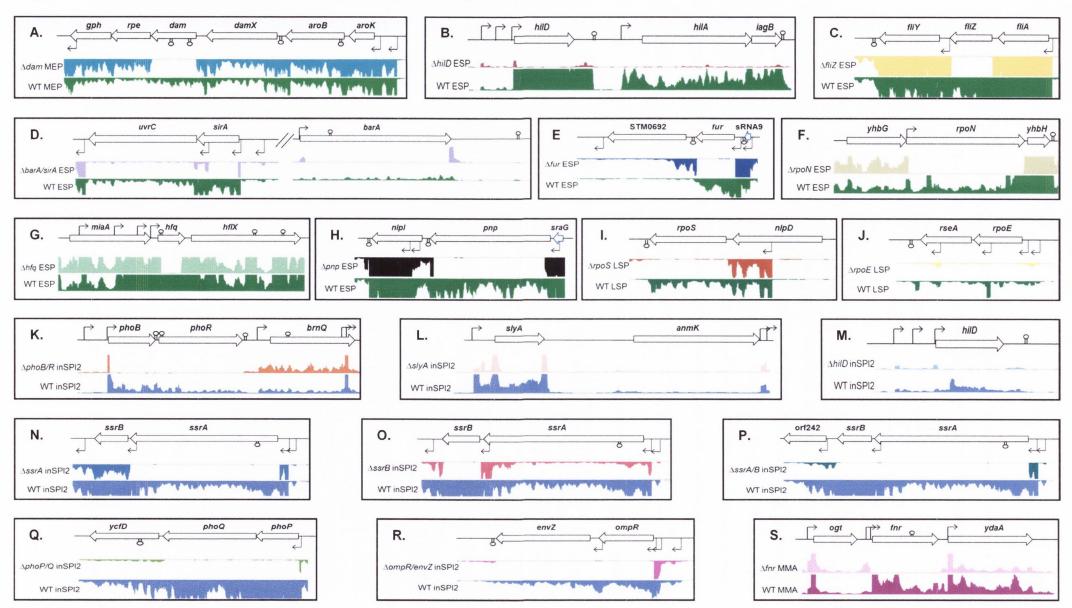
ycfD encodes a gene of unknown function and is transcribed from the phoP/Q promoter (Figure 3.4 Q). Expression of ycfD is reduced approximately 5-fold as a result of deleting the phoP and phoQ genes. Transcription from the phoP/Q promoter is auto-regulated by PhoP, so a reduction in ycfD expression would be expected in the absence of PhoP. However, deletion of the ycfD gene does not seem to have a broad effect on cellular regulons as a kanamycin resistance cassette was previously introduced into this gene to

mediate co-transduction of the constitutive phoQ24 mutation into various genetic backgrounds by other laboratories (Golubeva et al., 2012).

The ydaA gene is encoded from its own promoter, downstream of the fnr gene. Expression of ydaA is reduced approximately 2-fold in the absence of FNR. YdaA is annotated as a putative universal stress protein and expression of ydaA is highest under anaerobic shock and during anaerobic growth (Kröger et al., 2013). It is, therefore, possible that the ydaA gene is regulated by FNR under anaerobic conditions, resulting in the decrease in ydaA expression in the  $\Delta fnr$  mutant. In E coli there is a regulatory link between FNR and ydaA via GadX (Constantinidou et al., 2006, Hodges et al., 2010). There is no gadX orthologue in S. Typhimurium however, but it may be possible that an alternative regulatory link exists between FNR and ydaA in S. Typhimurium. It can also not be ruled out that the deletion of the fnr gene has a slight transcriptional polar effect on ydaA expression although the kanamycin resistance cassette was removed from the  $\Delta fnr$  mutant strain.

Overall, the recombinase-mediated deletion of 18 genes resulted in the successful construction of a panel of mostly non-polar virulence-associated regulatory mutants. Figure 3.4 demonstrates that each gene is transcribed under the conditions used for RNA-seq based investigation of the regulon of that protein. Figure 3.4 also demonstrates the specificity of the RNA-seq technology used in this study as sequencing reads did not map to the deleted region of each gene.

Figure 3.4 Confirmation of 19 chromosomal deletions by RNA-seq



#### Figure 3.4 Confirmation of 19 chromosomal deletions by RNA-seq

A-S. Integrated Genome Browser (IGB) visualisation of sequence reads for each deleted gene and the surrounding genes in the relevant mutant strain and wild-type comparator, grown under the same conditions. The colours of each track represent the sequencing reads which map to that locus and the height of reads is directly proportional to the level of expression at that locus (Nicol *et al.*, 2009). These panels demonstrate that the gene encoding each regulator is expressed in wild-type under the growth condition chosen for analysis of that regulator, and that no sequencing reads map to the deleted region in the mutant strain. Neighbouring genes are generally not affected by polar mutations (see text for exceptions). White arrows with a black outline denote protein coding genes. White arrows with a blue outline denote sRNA genes. Black bent arrows indicate TSS. All arrows indicate the direction of transcription. Predicted rho ( $\rho$ )-independent terminators (Kröger et al., 2012) are denoted by stem-loop structures.

#### 3.2.3 Identification of previously determined regulons validates the RNA-seq approach

The use of RNA-seq is a relatively new method of determining the regulon of a specific TF. In the past, transcriptional regulation studies focusing on a few specific genes were conducted using transcriptional fusions, northern blots or qRT-PCR. More recently, larger scale transcriptomic studies were undertaken using DNA microarrays which greatly improved our understanding of transcriptional regulation. However, there are some drawbacks to using microarrays for transcriptomic analysis. The resolution of microarrays can be quite low while background can be quite high, due to non-specific hybridisation between cDNA and probes (Aikawa *et al.*, 2010). Moreover, the dynamic range of microarrays is limited by the use of fluorescently-labelled DNA, detection of which can be saturated and subtle changes in gene expression may not be detected. RNA-seq, on the other hand, allows all transcription to be studied in an unbiased manner without prior knowledge of the DNA sequences that are being transcribed and with no limits on the dynamic range (Croucher & Thomson, 2010). For this reason, RNA-seq is the ideal method to use for identifying new small RNAs and the regulation of their expression, as well as to expand our knowledge of the complexity of bacterial regulons.

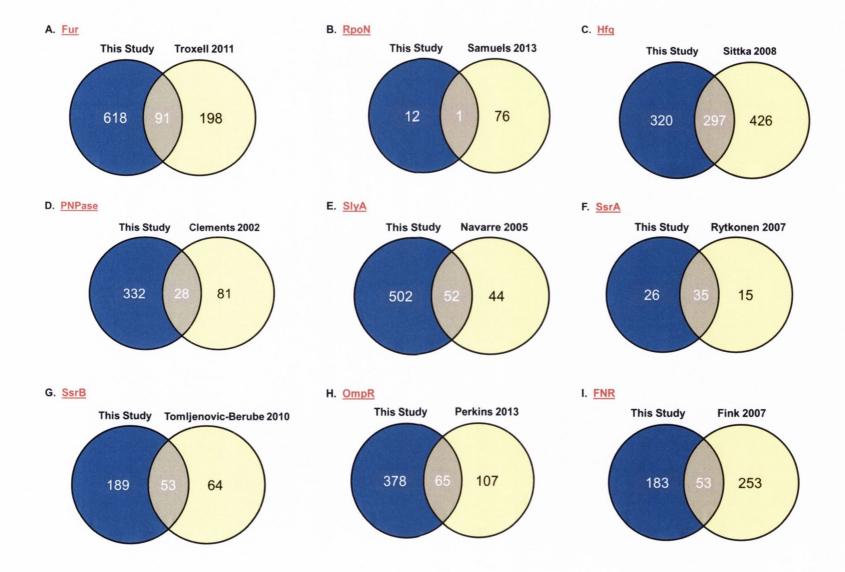
The protein coding gene targets of many *Salmonella* TFs have been well studied using transcriptional fusions and microarrays (Balbontin et al., 2006, Fink et al., 2007, Troxell *et al.*, 2011a, Thijs *et al.*, 2007, Tomljenovic-Berube et al., 2010, Navarre et al., 2005, Rytkonen *et al.*, 2007, Samuels *et al.*, 2013, Yoon *et al.*, 2009, Perkins et al., 2013, Clements et al., 2002, Ygberg et al., 2006). The *S.* Typhimurium regulons that have been

defined using microarray-based transcriptomics have recently been reviewed by our laboratory (Hebrard et al., 2011). These important studies have provided a useful platform for us to validate the use of RNA-seq as a method of investigating bacterial regulons. This validation gives us confidence that RNA-seq-based transcriptomics will prove to be useful in the characterisation of how sRNAs are regulated and the role played by sRNAs within important cellular regulons.

Figure 3.5 (A-I) demonstrates the similarities and differences in a selection of important *S*. Typhimurium regulons as determined by RNA-seq ("This Study") and previously published microarray-based studies. These comparisons are discussed in sections 3.2.3.1 through 3.2.3.9. All gene lists are included in Appendix IV.

Figure 3.5 J shows the IGB visualisation of RNA-seq data from the SPI1 locus in 15 regulatory mutants and wild-type comparators. The SPI1 locus is not expressed during logarithmic growth in Lennox medium and the absence of Dam-mediated methylation does not affect SPI1 gene expression under these growth conditions. As expected, SPI1 genes are highly expressed at ESP (SPI1-inducing conditions) and are not expressed in the absence of HilD. Expression of the SPI1 genes is reduced in the absence of the SPI1-associated regulators, FliZ, BarA/SirA and Fur. These effects are mediated through HilD, at the post-transcriptional level, as previously described (Figure 1.3) and as expected, transcription of hilD is not significantly affected in the SPI1-associated regulatory mutants. The SPI1 locus is not expressed at LSP or under SPI2-inducing conditions. Expression of hilD and the other SPI1 genes is de-repressed in many of the SPI2-associated regulators under SPI2-inducing conditions, however. This de-repression is currently the subject of further investigation, and appears to serve as a mechanism of switching-off invasion-associated genes when S. Typhimurium is in the intracellular environment. PhoP repression of SPI1 via hilA has previously been reported (Behlau & Miller, 1993, Bajaj et al., 1996). SPI1 gene expression is not significantly affected in the absence of FNR under the conditions used in this study. This will be discussed in section 3.2.3.9.

Figure 3.5 Validation of RNA-seq of mutant strains as a method of investigating Salmonella regulons



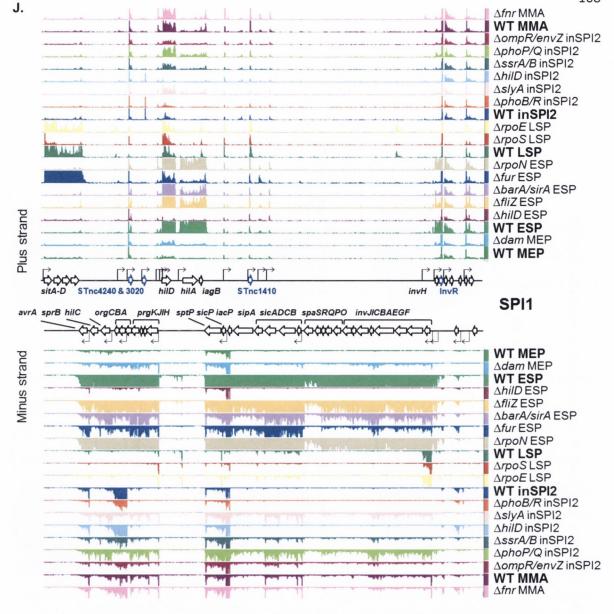


Figure 3.5 Validation of RNA-seq of mutant strains as a method of investigating Salmonella regulons

**A-I.** Venn diagrams comparing previously published regulons, derived from microarray data, with the RNA-seq-based regulons generated in this study. With the exception of the RpoN regulon (B) there is generally a significant of overlap between the previously published and new regulons. Differences in the amounts of differentially expressed genes in a regulon determined by the two techniques typically reflect strain and growth differences, as well as the increased sensitivity and dynamic range of RNA-seq-based technology compared to microarrays. J. IGB visualisation of the SPI1 pathogenicity island, following transcriptomic analysis of 15 regulatory mutants and wild-type comparators. The colours of each track represent the sequencing reads which map to that locus and are directly proportional to the level of expression at that locus. annotation tracks arrows with a black outline denote protein coding genes and protein-coding gene names are labelled in black. White arrows with a blue outline denote sRNA genes and sRNA gene names are labelled in blue. Black bent arrows indicate TSS. All arrows indicate the direction of transcription, + and - strands are indicated. Strain names and growth conditions are labelled at the right of the image. Wild-type comparators are labelled in bold.

#### 3.2.3.1 The Fur regulon

Figure 3.5 A shows the protein coding and non-coding genes which demonstrate a 2.5-fold or greater change in expression in an isogenic mutant of Fur, relative to wild-type, in this RNA-seq-based study and following microarray-based transcriptomic profiling of the  $\Delta fur$  mutant (Troxell et al., 2011a). There are 91 genes which are common to both datasets, including up-regulated genes (repressed by Fur) which encode iron transport proteins (sitABC, fhuF, feoB) and siderophore-associated proteins (iroBCDEN, fes, fepA, entABCE) as well as down-regulated genes (activated by Fur) which encode the superoxide dismutase (sodB), the fumarate reductase operon (frdABD)and proteins involved in anaerobic metabolism (*dmsA*). Differences in growth conditions between the two studies could account for the large numbers of genes which are differentially expressed in one study but not in the other. Troxell et al determined the Fur regulon under fully anaerobic conditions. Wild-type and the  $\Delta fur$  mutant were cultured to OD<sub>600</sub>0.3 in buffered Luria-Bertani containing MOPS and xylose (LB-MOPS-X) and anaerobic conditions were maintained using an anaerobic gas chamber. The growth under anaerobic conditions led to differential expression of a number of genes involved in anaerobic metabolism such as, the cytochrome C oxidase cyo operon, the eut operon which is necessary for growth using ethanolamine as an alternative carbon source and is thought to be important under the low oxygen conditions during Salmonella infection (Garsin, 2010), the tdc operon which is responsible for the anaerobic transport and metabolism of L-threonine and L-serine and the glp operon which is necessary for the fermentation of glycerol under anaerobic conditions. A number of these anaerobically induced operons are not expressed under the conditions used in our study and were removed during the analysis of the RNA-seq data.

In order to focus on the role of Fur in regulation of SPI1 and SPI1-associated genes the wild-type and  $\Delta fur$  mutant strains used in this study were grown under SPI1-inducing conditions (ESP). The 618 genes which are only differentially expressed in this study comprise SPI1 genes, SPI4 genes and flagellar genes. A link between SPI1 and Fur regulation has been previously reported (Ellermeier & Slauch, 2008, Troxell et al., 2011b), and while no direct link between Fur and flagellar genes has yet been reported, SPI1 and flagellar genes are co-regulated (Thijs et al., 2007, Singer *et al.*, 2014, Ellermeier & Slauch, 2003). SPI1 and SPI4 genes are also co-regulated (Main-Hester *et al.*, 2008). The RNA-seq dataset also contains 78 differentially expressed sRNAs, many

of which were only recently annotated (Kröger et al., 2013) and would not have been included on the microarray used by Troxell *et al*. In addition, the improved sensitivity and dynamic range of RNA-seq compared to microarrays likely detects more subtle changes in gene expression (Ozsolak & Milos, 2011).

## 3.2.3.2 The RpoN/ $\sigma^{54}$ regulon

In Figure 3.5 B the overlap between the RNA-seq-based analysis of wild-type and an isogenic mutant of RpoN with microarray data, which also analysed wild-type and  $\Delta rpoN$ strains (Samuels et al., 2013), is minimal. The compared gene lists comprise genes which exhibit at least a 3-fold decrease in expression in the  $\Delta rpoN$  mutant compared to wild-type in both studies. Only one gene, pspA which encodes a bacteriophage shock protein, is common to both datasets. In this case the Samuels et al dataset likely represents the more comprehensive  $\sigma^{54}$  regulon as it involved the use of a promiscuous bacterial enhancer binding protein. As previously discussed,  $\sigma^{54}$  requires a bacterial enhancer-binding protein (bEBPs) and hydrolysis of ATP to form a transcriptionally competent open complex. Specific environmental stimuli are required to activate bEBPs and so analysis of  $\sigma^{54}$ -dependent promoters under one environmental condition may not detect transcription from promoters whose bEBP is not active under that condition. Samuels et al used a plasmid-borne constitutively expressed bEBP from Sinorhizobium meliloti, DctD, to ensure transcription from all  $\sigma^{54}$ -dependent promoters. The pspA promoter is the only predicted  $\sigma^{54}$ -dependent promoter that was detected when mapping TSS in S. Typhimurium under the conditions used in this study (ESP) (Kröger et al., 2012). In addition to this key difference between the two studies, the numbers of differentially expressed genes in the  $\Delta rpoN$  mutant could be due to a  $\sigma^{70}$  promoter overlapping the  $\sigma^{54}$  promoter, as has previously been shown for the sRNAs glmY and glmZ which are transcribed from a  $\sigma^{54}$  promoter but their expression is not decreased in a  $\Delta rpoN$  mutant due to activation of transcription by  $\sigma^{70}$  (Gopel et al., 2011). However, some genes such as the nar and nir operons which are involved in nitrate metabolism and transport are differentially expressed in the RNA-seq dataset and are likely to be direct targets of the nitrogen-assimilation  $\sigma$ -factor,  $\sigma^{54}$ .

#### 3.2.3.3 The Hfq regulon

Figure 3.5 C shows a comparison of the genes which were at least 2-fold differentially expressed in an isogenic mutant of the RNA chaperone Hfq, compared to the wild-type strain grown to ESP, by RNA-seq in this study and by microarray (Sittka et al., 2008). sRNA genes were omitted from this analysis due to the large numbers of sRNA genes which are likely to be affected by the absence of Hfq. Apart from the method of transcriptome profiling, the main difference between the two studies is the strain background used. In this study 4/74 was used, while the histidine auxotroph derived from 4/74 (SL1344) was used by Sittka et al. There is a large overlap of 297 genes that are differentially expressed in both studies, however there are also 320 genes and 426 genes that are only differentially expressed in this study and in the Sittka et al study, respectively. A number of the genes which are only differentially expressed in the Sittka et al study are genes of unknown function or are SPI2 effectors which have low levels of expression or are not expressed at all and were filtered out of the RNA-seq dataset during analysis. Other genes in the Sittka et al dataset are excluded from the RNA-seq dataset as they were slightly under the 2-fold cut-off. The increased dynamic range and sensitivity of the RNA-seq technology allows for a more refined measurement of gene expression levels. Similarly, a number of the genes which are only differentially expressed in the RNA-seq dataset are just above the 2-fold cut-off and the more subtle change in expression may not have been detected using the microarray-based method.

#### 3.2.3.4 The PNPase regulon

In Figure 3.5 D, genes which show at least a 2-fold change in expression between wild-type and a deletion mutant of the exoribonuclease PNPase by RNA-seq were compared to genes which demonstrate a 2-fold, or greater, change in expression as determined by microarray analysis of a *pnp* mutant. The *pnp* mutant used in the microarray-based study contained a point mutation in the *pnp* gene, which results in a truncated PNPase protein without exonuclease activity (Clements et al., 2002). There is an overlap of 28 genes between both datasets. The genes which are differentially expressed in both studies mainly comprise genes that encode components of the SPI2 TTSS apparatus and SPI2 effector proteins. Clements *et al* reported increased levels of SPI2 gene transcripts, and a concomitant increase in bacterial growth rates within murine

macrophages and increased persistence of a pnp mutant in infected mice, highlighting the importance of PNPase as a virulence factor. Clements et al also reported increased levels of SPI1 gene transcripts and increased stability of the SPI1 effector sipC. However, only slight increases (below the 2-fold cut-off) in SPI1 gene expression levels were detected by RNA-seq in this study. This may be due to the difference in growth conditions between the studies. Clements et al quantified gene expression in rich media at OD<sub>600</sub>1.0 (LEP) while RNA for analysis by RNA-seq in this study was extracted from bacteria grown in rich media to OD<sub>600</sub>2.0 (ESP). As ESP is the optimal condition for the expression of SPI1 genes the gene transcripts are likely to be very stable under this condition and may not be subject to degradation by PNPase. A transcriptomic-based study of a pnp mutant grown under macrophage-mimicking conditions demonstrated a decrease in SPI1 gene expression in the absence of PNPase (Ygberg et al., 2006), highlighting that PNPase can play a differential role in the regulation of virulence gene expression under different environmental conditions. A number of genes which are only differentially expressed in this study includes small RNA genes which may be targets for PNPase-mediated degradation, but these genes were not included on the 2002 microarray. In addition, the microarray used by Clements et al featured only 71% of all the protein coding sequences annotated in the Salmonella LT2 genome at the time. protein-coding genes which were found to be differentially expressed by RNA-seq, 24% were not included on the microarray.

#### 3.2.3.5 The SlyA regulon

Genes showing a 2-fold or greater change in expression following RNA-seq analysis of wild-type and a  $\Delta slyA$  mutant grown under the same conditions were compared to genes showing a 2-fold or greater change in expression following microarray analysis of wild-type and a  $\Delta slyA$  mutant (Navarre et al., 2005) (Figure 3.5 E). Genes which are differentially expressed in both datasets mostly include SPI2 regulator and effector genes as well as genes which have been shown to be co-regulated by PhoP and SlyA. The transcriptome profiling from both studies was conducted under different conditions, which accounts for the genes which are differentially expressed in one study but not the other. In this study the SlyA regulon was investigated under SPI2-inducing conditions (PCN minimal medium, pH 5.8) (Lober et al., 2006), while Navarre *et al* used RNA from

strains grown to ESP in rich media. The differentially expressed genes in this study alone include a number of genes which encode SPI1 regulators and effectors, as well as flagellar proteins, which are normally repressed under SPI2-inducing conditions, but are de-repressed in a  $\Delta slyA$  mutant under SPI2-inducing conditions (Figure 3.5 J). This de-repressive effect in the absence of SlyA is not evident at ESP when SPI1 genes are already highly expressed. The majority of genes which are only differentially expressed in the Navarre *et al* study are not expressed under SPI2-inducing conditions and were filtered out of the analysis of the SlyA regulon from the RNA-seq data.

#### 3.2.3.6 The SsrA regulon

Figure 3.5 F shows the genes showing the largest fold-decrease (at least 8-fold) following RNA-seq analysis of wild-type and  $\triangle ssrA$  strains, compared to the 50 genes showing the largest fold-decrease between the two strains as analysed by microarray (Rytkonen et al., Thirty-five genes encoding structural components, chaperones and effector proteins of the SPI2 TTSS have significantly decreased expression in both studies. The additional number of SPI2-associated genes which are differentially expressed in this study but not in Rytkonen et al may highlight the increased sensitivity of the RNA-seq technology in comparison to the microarray-based approach. RNA-seq also identified 2 sRNAs which belong to the SsrA regulon which were not detected by microarray. Some of the genes which are only differentially expressed in the study by Rytkonen et al are not expressed under the conditions used in this study and so were filtered out during data analysis. Other genes may be differentially expressed in the Rytkonen et al study and not in this study due to differences in growth conditions between the two studies. Rytkonen et al extracted RNA from overnight cultures grown in MgM-Mes at pH 4.5 while cultures from mid-log growth in SPI2-inducing PCN media pH 5.8 were used in this study. In addition to the more acidic nature of the medium and different growth phases of the cultures, MgM-Mes contains a higher level of phosphate but lower levels of magnesium than PCN medium (Beuzon et al., 1999). The differentially expressed genes encode elements of the TCA cycle, the alternative sigma factor  $\sigma^{S}$  and cold shock proteins, which reflects the more stressful conditions used in the Rytkonen et al study.

#### 3.2.3.7 The SsrB regulon

In Figure 3.5 G the most significantly down-regulated (2.5-fold or greater) genes following RNA-seq of wild-type and a  $\triangle ssrB$  mutant were compared to the genes showing the largest fold-decrease in expression following microarray analysis of the SsrB regulon (Tomljenovic-Berube et al., 2010). Similarly to the previously mentioned SsrA regulon, the genes which are down-regulated in both datasets encode regulators and components of the SPI2 TTSS. A number of the genes which are only down-regulated in the Tomljenovic-Berube study are not expressed under the conditions tested in this study and so were filtered out of the RNA-seq data analysis. Many of the genes that were identified as only being differentially expressed in this study are at the lower levels of absolute expression and differential expression between wild-type and  $\triangle ssrB$ , and so may not have been detected by the less sensitive microarray technology. Genes which are only differentially expressed in this study include phoP, the response regulator of the PhoP/Q TCS, the expression of which decreases 3-fold in the  $\triangle ssrB$  mutant. Therefore many of the genes which are only down-regulated in the RNA-seq dataset are PhoP-dependent genes such as pag genes and ugtL. The effects of deleting the ssrB gene on PhoP-dependent genes are unlikely to be direct, however the possibility of inverse regulation of phoP expression by SsrB had not been previously reported and highlights how the increased sensitivity of the RNA-seq technology can detect interactions which may not have been previously identified.

#### 3.2.3.8 The OmpR regulon

Perkins *et al* investigated differential gene expression between wild-type and  $\Delta ompR$  strains grown to OD<sub>600</sub>0.6 in LB medium (Perkins et al., 2013), while mid-log (OD<sub>600</sub>0.3) cultures grown in SPI2-inducing PCN medium, in order to focus on the SPI2-association of the OmpR/EnvZ TCS, were used in this study (Figure 3.5 H). Genes were considered to be differentially expressed if they demonstrated a 2-fold, or greater, change in expression in either study. Despite the differences in culture conditions between the two studies, the overlap of genes which are differentially expressed in both studies is quite high. The majority of genes which are up-regulated in the absence of OmpR in both studies include SPI1-associated and flagellar genes. The conditions tested in both studies are not optimal for SPI1 expression and the up-regulation of these genes in  $\Delta ompR$ 

mutants highlights the role played by OmpR in repressing SPI1. OmpR binding to and repressing expression from the *hilD* promoter when grown in LB has been previously reported (Cameron & Dorman, 2012), and OmpR binding to the *hilD* promoter region increases under acidic conditions (Quinn et al., 2014), similar to the more acidic conditions used in this study. Genes encoding SPI2 regulators, components of the TTSS apparatus and effector proteins comprise the down-regulated genes in both datasets. A number of the genes which are differentially expressed in one study and not the other reflect the differences in growth conditions between the two studies and the different lifestyles of bacteria growing logarithmically in rich medium, or in minimal macrophage-mimicking media. Many of the genes which were found to be differentially expressed by Perkins *et al* were not expressed under the conditions tested in this study and were filtered out during analysis. Additionally, genes which are only differentially expressed in this study include a number of chemotaxis and motility genes, as well as additional SPI2 genes, and these interactions were not detected using the less sensitive microarray-based approach.

#### 3.2.3.9 The FNR regulon

Figure 3.5 I shows the FNR regulon under micro-aerobic conditions, as analysed by RNA-seq, compared to the FNR regulon under anaerobic conditions as determined by microarray (Fink et al., 2007). RNA extracts from cultures grown to mid-log phase in LB-MOPS-X pH 7.4 in an anaerobic gas chamber were used in the microarray-based study. RNA was extracted for RNA-seq at OD<sub>600</sub>0.3 from statically grown cultures in MMA medium supplemented with TMAO and fumarate as alternative electron acceptors, pH 7 (as described in Table 2.1). Only genes showing a 2.5-fold, or greater, change in expression in the  $\Delta fnr$  mutant compared to the wild-type strain were included in the comparison between the two studies. The majority of the genes which are down-regulated in both studies comprise genes which encode components of flagella biosynthesis, motility and chemotaxis regulons. The fnr mutant of S. Typhimurium lacks flagella and is non-motile (Fink et al., 2007). Genes involved in anaerobic metabolism, such as dmsABC, dcuB and fumB are also included in both data sets. Twenty-five sRNAs are included in the number of differentially expressed genes identified by RNA-seq in this study, and these sRNA genes were not detected by microarray analysis. Included in this group is the sRNA, FnrS, which is an anaerobically-inducible, FNR-dependent sRNA in *E. coli* (Boysen et al., 2010, Durand & Storz, 2010), but regulation of the FnrS orthologue in *S.* Typhimurium has not been previously reported.

Another striking difference between the two datasets is the increased expression of genes encoding SPI2 regulators, components of the SPI2 TTSS apparatus and SPI2 effector proteins in this study, while these genes are not up-regulated in the absence of FNR in the study by Fink et al, with the exception of sopD2 and sseL which are common to both datasets. Fink et al did not detect differential gene expression of the majority of SPI2 genes, however an fnr mutant is attenuated for survival and replication within macrophages. This would be consistent with FNR modulation of SPI2 gene expression, but may not have been detected under the conditions used in the microarray-based study. Many of the genes which are only differentially expressed in the Fink et al study were not expressed in the conditions used in this study and were, therefore, filtered out of the analysis. Genes encoding SPI1 regulators, components of the SPI1 TTSS and SPI1 effector proteins are only down-regulated in the microarray-based study. It is unclear from the Fink et al study whether SPI1-associated genes are directly regulated by FNR or if the FNR-dependence of these genes is due to FNR-mediated down-regulation of the genes encoding FliZ and FliA, which play roles in the regulation of SPI1 (Lucas et al., 2000). Another explanation may be that the disruption in anaerobic metabolism in the absence of FNR leads to induction of hilA via BarA/SirA, due to accumulation of fatty acid metabolites, such as acetate (Lawhon et al., 2002). The differences in oxygen levels and thus in differential anaerobic metabolism gene expression may account for the variation in SPI1 gene expression between the two studies.

# 3.2.4 Validation of the RNA-seq approach for detecting regulatory inputs to sRNA gene expression

Typically, transcriptional regulation of sRNAs by TFs and  $\sigma$ -factors has been determined using northern blotting and transcriptional fusions. Recently an RNA-seq-based method was used to expand the regulon of  $\sigma^{38}$  and to identify  $\sigma^{38}$ -dependent sRNAs (Levi-Meyrueis *et al.*, 2014). cDNA library preparation for the study by Levi-Meyrueis *et al.* involved ribosomal RNA depletion and enrichment for sRNAs, while library preparation

for this study did not include any depletion or enrichment methods prior to sequencing. To ensure that our RNA-seq-based method would accurately report differential sRNA gene expression, previously published regulatory interactions between TFs and sRNAs were compared to the TPM values for the same TFs and sRNAs from this RNA-seq-based study (Table 3.1). This analysis focuses on traditional transcriptional activation or repression rather than regulation of sRNA activity or stability, therefore isogenic mutants of Hfq and PNPase were omitted from this analysis and both mutant datasets will be analysed separately (section 3.2.12 and 3.2.13).

Over 91% of the previously published regulatory interactions between TFs or  $\sigma$ -factors and sRNA genes are detected by our RNA-seq-based method in S. Typhimurium, highlighting the reliability of the technique as a discovery tool to identify further regulatory inputs to sRNA gene expression. Four sRNA genes (csrC, gcvB, glmY and sroC) do not show the expected differential expression in an rpoS mutant reported by Levi-Meyrueis et al. The differences in gene expression between these 4 sRNA genes may be due to strain differences as the Levi-Meyrueis et al study used S. Typhimurium 14028 rather than 4/74. Growth conditions between the two studies were also different as Levi-Meyrueis et al used RNA from late stationary phase cultures grown for 18 hours while our LSP cultures are grown to OD<sub>600</sub>2.0 + 6 hours (approximately 12 hours in total), and it is possible that regulation of these sRNAs by  $\sigma^{38}$  is more important at a later stage of stationary phase. The antisense transcript AmgR was not expressed under the conditions tested in this study and thus differential expression in a  $\Delta phoP$  mutant could not be detected (Lee & Groisman, 2010). The sRNAs GlmY and GlmZ are transcribed from  $\sigma^{54}$ -dependent promoters however, as previously discussed, deletion of rpoN does not result in a reduction in expression of the glmY and glmZ genes due to an overlapping  $\sigma^{70}$ -dependent promoter. In fact a slight increase in glmZ expression in the absence of  $\sigma^{54}$ has also been reported (Gopel et al., 2011).

Transcriptional regulation of the *mgrR*, *omrA*, *omrB*, *micF*, *fnrS* and *micL* genes by transcription factors, as shown in Table 3.1, was demonstrated in *E. coli* but not in *Salmonella*. Our RNA-seq data show that the *S*. Typhimurium orthologous sRNA genes are regulated in the same way as in *E. coli*.

Table 3.1 Comparison of RNA-seq-based transcriptomic data with published regulatory interactions between TFs and sRNAs

sRNA	Published	Published	Reference	WT TPM <sup>a</sup>	mutant TPM <sup>b</sup>	Fold Change
	Regulator	Interaction		(RNA-seq)	(RNA-seq)	(RNA-seq)
	on determined in Saln	nonella				
Typhimurium InvR	HilD	activation	(Desiffer et al. 2007)	32,390	454	71-fold decrease
DapZ	HilD	activation	(Pfeiffer et al., 2007) (Chao et al., 2012)	31,651	1,989	16-fold decrease
CsrB	BarA/SirA	activation	(Chao et al., 2012) (Lawhon et al., 2002, Teplitski et al., 2003)	8,117	1,186	7-fold decrease
CsrC	BarA/SirA	activation	(Fortune et al., 2006)	15,043	4,421	3-fold decrease
RyhB-1	Fur	repression	(Padalon-Brauch et al., 2008)	32	18,627	582-fold increase
GlmY	$\sigma^{\scriptscriptstyle  m N}$	$\sigma^N$ -dependent	(Gopel et al., 2011)	4,104	3,153	1-fold decrease
GlmZ	$\sigma^{\scriptscriptstyle  m N}$	$\sigma^N$ -dependent	(Gopel et al., 2011)	3,040	3,316	1-fold decrease
SdsR	$\sigma^{s}$	activation	(Frohlich et al., 2012)	182,917	285	642-fold decrease
SraL	$\sigma^{\scriptscriptstyle S}$	activation	(Silva et al., 2013)	2,437	34	72-fold decrease
CsrC	$\sigma^{s}$	activation	(Levi-Meyrueis et al., 2014)	3,851	2,925	1-fold-decrease
CyaR	$\sigma^{\scriptscriptstyle S}$	repression	(Levi-Meyrueis et al., 2014)	823	3,980	5-fold increase
DsrA	$\sigma^{\scriptscriptstyle S}$	repression	(Levi-Meyrueis et al., 2014)	958	4,259	4-fold increase
GcvB	$\sigma^{\scriptscriptstyle S}$	activation	(Levi-Meyrueis et al., 2014)	155	325	2-fold increase
GlmY	$\sigma^{s}$	activation	(Levi-Meyrueis et al., 2014)	2,036	2,386	1-fold increase
IsrI	$\sigma^{s}$	activation	(Levi-Meyrueis et al., 2014)	27	10	3-fold decrease
IstR-1,2	$\sigma^{\scriptscriptstyle S}$	activation	(Levi-Meyrueis et al., 2014)	253	43	6-fold decrease
MicA	$\sigma^{\scriptscriptstyle S}$	activation	(Levi-Meyrueis et al., 2014)	12,736	5,082	3-fold decrease
OmrA	$\sigma^{s}$	activation	(Levi-Meyrueis et al., 2014)	2,418	220	11-fold decrease
OxyS	$\sigma^{\scriptscriptstyle S}$	activation	(Levi-Meyrueis et al., 2014)	96	37	3-fold decrease
RybA	$\sigma^{\scriptscriptstyle S}$	activation	(Levi-Meyrueis et al., 2014)	2,325	513	5-fold decrease
RybB	$\sigma^{\scriptscriptstyle S}$	activation	(Levi-Meyrueis et al., 2014)	4,780	1,736	3-fold decrease
RybD	$\Box^{S}$	repression	(Levi-Meyrueis et al., 2014)	280	713	3-fold increase

Table 3.1 Comparison of RNA-seq-based transcriptomic data with published regulatory interactions between TFs and sRNAs

RydC	$\sigma^{\rm S}$	repression	(Levi-Meyrueis et al., 2014)	9,448	51,221	5-fold increase
RyeF	$\sigma^{\scriptscriptstyle S}$	activation	(Levi-Meyrueis et al., 2014)	255	142	2-fold decrease
RygC	$\sigma^{\scriptscriptstyle S}$	activation	(Levi-Meyrueis et al., 2014)	4,218	1,302	3-fold decrease
RygD	$\sigma^{\scriptscriptstyle S}$	activation	(Levi-Meyrueis et al., 2014)	4,611	1,681	3-fold decrease
RyhB-1	$\sigma^{\scriptscriptstyle S}$	activation	(Levi-Meyrueis et al., 2014)	3,742	234	16-fold decrease
RyhB-2	$\sigma^{s}$	activation	(Levi-Meyrueis et al., 2014)	9,267	845	11-fold decrease
SraC	$\sigma^{s}$	repression	(Levi-Meyrueis et al., 2014)	115	434	4-fold increase
sRNA10	$\sigma^{\scriptscriptstyle S}$	activation	(Levi-Meyrueis et al., 2014)	487	85	6-fold decrease
SroC	$\sigma^{\scriptscriptstyle S}$	activation	(Levi-Meyrueis et al., 2014)	4,721	4,149	1-fold decrease
STnc1060	$\sigma^{\scriptscriptstyle S}$	activation	(Levi-Meyrueis et al., 2014)	2,065	539	4-fold decrease
STnc1080	$\sigma^{\scriptscriptstyle S}$	repression	(Levi-Meyrueis et al., 2014)	176	508	3-fold increase
STnc1110	$\sigma^{\scriptscriptstyle S}$	activation	(Levi-Meyrueis et al., 2014)	911	34	27-fold decrease
STnc1200	$\sigma^{\scriptscriptstyle S}$	repression	(Levi-Meyrueis et al., 2014)	142	1,150	8-fold increase
STnc1220	$\sigma^{\scriptscriptstyle S}$	activation	(Levi-Meyrueis et al., 2014)	246	22	11-fold decrease
STnc1280	$\sigma^{\scriptscriptstyle S}$	repression	(Levi-Meyrueis et al., 2014)	4,204	9,579	2-fold increase
STnc1300	$\sigma^{\scriptscriptstyle S}$	repression	(Levi-Meyrueis et al., 2014)	486	1,455	3-fold increase
STnc1330	$\sigma^{\scriptscriptstyle S}$	activation	(Levi-Meyrueis et al., 2014)	759	12	63-fold decrease
STnc1380	$\sigma^{\scriptscriptstyle S}$	repression	(Levi-Meyrueis et al., 2014)	19	56	3-fold increase
STnc1390	$\sigma^{\scriptscriptstyle S}$	activation	(Levi-Meyrueis et al., 2014)	1,312	340	4-fold decrease
STnc150	$\sigma^{\scriptscriptstyle S}$	activation	(Levi-Meyrueis et al., 2014)	68	14	5-fold decrease
STnc1560	$\sigma^{\scriptscriptstyle S}$	activation	(Levi-Meyrueis et al., 2014)	455	83	5-fold decrease
STnc2080	$\sigma^{\scriptscriptstyle S}$	activation	(Levi-Meyrueis et al., 2014)	10,625	553	19-fold decrease
STnc290	$\sigma^{s}$	repression	(Levi-Meyrueis et al., 2014)	849	3,270	4-fold increase
STnc540	$\sigma^{s}$	activation	(Levi-Meyrueis et al., 2014)	361	73	5-fold decrease
STnc580 (FnrS)	$\sigma^{\rm S}$	repression	(Levi-Meyrueis et al., 2014)	15	37	2-fold increase

Table 3.1 Comparison of RNA-seq-based transcriptomic data with published regulatory interactions between TFs and sRNAs

STnc750	$\sigma^{\mathrm{s}}$	activation	activation (Levi-Meyrueis et al., 2014)		104	7-fold decrease
RybB	$\sigma^{\!\scriptscriptstyle E}$	activation	(Papenfort et al., 2006)	4,780	57	84-fold decrease
MicA	$\sigma^{\scriptscriptstyle E}$	activation	(Papenfort et al., 2006)	12,736	90	142-fold decrease
AmgR	PhoP/Q	activation	(Lee & Groisman, 2010)	5	2	Not expressed <sup>c</sup>
MicC	OmpR/EnvZ	repression	(Pfeiffer et al., 2009)	148	714	5-fold increase
Regulatory interactio	on determined in Esche	erichia coli				
MgrR	PhoP/Q	activation	(Moon & Gottesman, 2009)	11,885	239	50-fold decrease
OmrA	OmpR/EnvZ	activation	(Guillier & Gottesman, 2006)	793	262	3-fold decrease
OmrB	OmpR/EnvZ	activation	(Guillier & Gottesman, 2006)	2,835	178	16-fold decrease
MicF	OmpR/EnvZ	activation	(Delihas & Forst, 2001)	453	147	3-fold decrease
FnrS	FNR	activation	(Boysen et al., 2010, Durand & Storz, 2010)	6,462	158	41-fold decrease
RyeF/MicL	$\sigma^{\!\scriptscriptstyle m E}$	activation	(Guo et al., 2014)	255	7	36-fold decrease

<sup>&</sup>lt;sup>a</sup> Wild-type comparator grown under same conditions as the specific isogenic mutant

<sup>&</sup>lt;sup>b</sup> Isogenic mutant that lacks the indicated regulator protein, grown under the same conditions as the wild-type comparator

<sup>&</sup>lt;sup>c</sup> Genes with TPM values < 10 are not expressed (as previously described in section 2.6.9)

# 3.2.5 Confirmation of the regulation of expression of new S. Typhimurium sRNAs

A total of 61% of all known or candidate S. Typhimurium sRNA genes are differentially expressed by a factor of 2.5-fold or greater in one or more of the 17 TF mutants ( $\Delta hfq$  and  $\Delta pnp$  excluded from analysis). In addition to the previously published regulatory interactions between TFs and sRNA genes, 44% of known or candidate S. Typhimurium sRNA genes were found to have novel putative regulatory interactions in this study. Northern blots were conducted to confirm some of these new regulatory interactions (Figure 3.6).

RNA-seq data from wild-type and a mutant strain lacking the primary SPI1 regulator, HilD, grown under SPI1-inducing conditions (ESP) resulted in a 17-fold decrease in expression of the sRNA STnc520 and this decrease in expression was validated by northern blot (Figure 3.6 A). This candidate regulatory interaction is discussed in detail in chapter 4.

Expression of the sRNA FnrS is dependent on FNR in *E. coli* (Boysen et al., 2010, Durand & Storz, 2010) and in *Salmonella* (Table 3.1). FnrS expression also decreases 32-fold in a  $\Delta fur$  mutant compared to the wild-type strain grown to ESP and this was validated by northern blot (Figure 3.6 B). Fur-mediated activation of the *fnrS* gene is likely to be indirect as Fur primarily acts as a gene repressor (Carpenter et al., 2009). Regulation of expression of the *fnrS* gene by Fur does not appear to be mediated via FNR, as expression of the *fnr* gene itself is increased slightly (1.6-fold) in the  $\Delta fur$  mutant. A strong overlap between the Fur and FNR regulons has previously been reported (Troxell et al., 2011a) and Fur-mediated co-regulation of FnrS may provide a method of cross-talk between the two regulons.

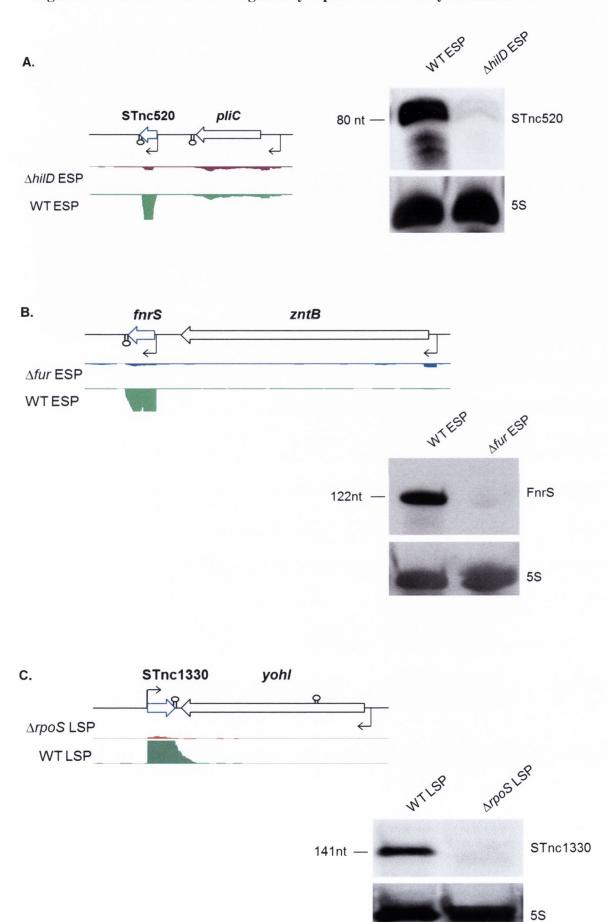
The dependence of STnc1330 expression on  $\sigma^{38}$  was previously reported but the interaction had not yet been validated by northern blot (Levi-Meyrueis et al., 2014). STnc1330 expression decreases 63-fold in the absence of  $\sigma^{38}$  in this study. While  $\sigma^{38}$  and the house-keeping sigma factor,  $\sigma^{70}$ , recognise identical -10 and -35 promoter consensus motifs, there are some promoter features which favour selection by  $\sigma^{38}$ , such as a conserved C at the -13 position relative to the TSS. There is also a preference for a C in the -8 position and an AT-rich sequence immediately downstream of the -10 site.  $\sigma^{38}$  also tends to operate more efficiently than  $\sigma^{70}$  at promoters which contain degenerate -35 sites

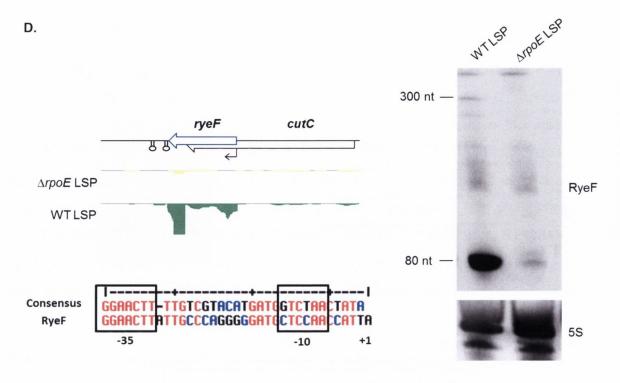
and at promoters that have sub-optimal spacer lengths between the -10 and -35 sites (Typas et al., 2007). These preferred nucleotides, a less conserved -35 box and a 16 bp spacer are all present in the STnc1330 promoter, suggesting selectivity for  $\sigma^{38}$ -dependent transcription from the STnc1330 promoter (see also Figure 3.9 G). Northern blot analysis (Figure 3.6 C) and further investigation of the regulation of STnc1330 expression by RpoS was carried out by Michael Beckett (BA mod. in Microbiology) as part of a supervised, undergraduate research project. STnc1330 expression was restored in a  $\Delta rpoS$  mutant by exogenous induction of RpoS expression. Furthermore, STnc1330 is most highly expressed under conditions known to activate members of the  $\sigma^{38}$  regulon such as, stationary phase growth, osmotic shock and carbon starvation (Kröger et al., 2013).

RpoE regulates transcription of the sRNAs RybB and MicA as part of the envelope stress response (Papenfort et al., 2006). RyeF is part of the RpoE regulon in *E. coli*, and was re-named MicL (Guo et al., 2014). RyeF accumulates in stationary phase (Kröger et al., 2013) and a *ryeF* mutant was identified in a screen of sRNAs and small proteins involved in resistance to cell envelope stress in *E. coli* (Hobbs *et al.*, 2010). The study by Guo *et al* found that MicL targets the outer membrane lipoprotein LPP. Our analysis of the RpoE regulon identified RyeF as a third sRNA in the RpoE regulon of *S.* Typhimurium. Expression of RyeF decreases 36-fold in the  $\Delta rpoE$  mutant compared to wild-type grown to LSP, while analysis of the *ryeF* promoter region identifies a consensus binding motif for  $\sigma^E$  (Figure 3.6 D). The -35 site is a perfect match for the  $\sigma^E$ -dependent promoter consensus sequence as previously determined (Skovierova *et al.*, 2006), while the -10 site only differs from the consensus sequence at 2 positions. Furthermore, an A at position +1 is very common in  $\sigma^E$ -dependent genes, and this A is present at the start of the *ryeF* gene.

The sRNA STnc1480 has multiple regulatory inputs from virulence-associated TFs under infection-relevant conditions. STnc1480 expression is reduced approximately 5-fold in mutants of SPI1-associated transcription factors, grown to ESP. However, STnc1480 is more highly expressed under conditions which mimic the macrophage environment (Kröger et al., 2013) and STnc1480 expression is up to 7-fold reduced in the absence of SsrA/B and OmpR/EnvZ, and up to 60-fold reduced in the absence of SlyA and PhoP. These regulatory inputs were validated by northern blot (Figure 3.6 E) and are discussed in detail in chapter 5.

Figure 3.6 Validation of new regulatory inputs on sRNAs by northern blot





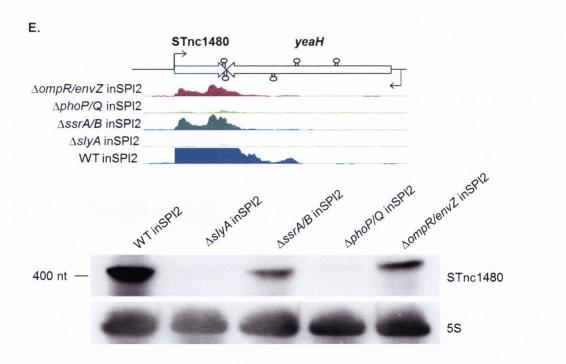


Figure 3.6 Validation of new regulatory inputs on sRNAs by northern blot

**A.** IGB visualisation and northern blot showing decreased expression of the STnc520 gene in  $\Delta hilD$  compared to wild-type at ESP. **B.** IGB visualisation and northern blot showing decreased expression of fnrS gene in  $\Delta fur$  compared to wild-type at ESP. **C.** IGB visualisation and northern blot showing decreased expression of the STnc1330 gene in  $\Delta rpoS$  compared to wild-type at LSP. **D.** IGB visualisation and northern blot showing decreased expression of the ryeF gene in  $\Delta rpoE$  compared to wild-type at LSP. Promoter alignment of ryeF promoter and consensus  $\sigma^E$  promoter. **E.** IGB visualisation and northern blot of STnc1480 with multiple regulatory inputs. Figure symbols are as described in Figure 3.4.

#### 3.2.6 Regulatory inputs on S. Typhimurium sRNA expression

Using the criteria defined in section 2.6.9, approximately 77% of all known or candidate sRNAs are expressed in wild-type 4/74 under all 5 environmental conditions used in this study, while a further 19% are expressed in wild-type under at least one environmental condition (Figure 3.7 A). Currently there are 280 known or candidate sRNA genes in S. Typhimurium 4/74 (Kröger et al., 2013), and as previously mentioned, 61% (170 genes) of all known or candidate sRNA genes are >2.5-fold differentially expressed in one or more of the 17 TF mutants. Only 1 sRNA gene, isrB-2, which is not expressed in wild-type under any of the tested environmental conditions is up-regulated and expressed in one of the mutant strains (>5-fold increase in expression in  $\Delta dam$  strain). Only 6 sRNA genes are not expressed in the wild-type strain grown under a particular environmental condition but are up-regulated and expressed in a TF mutant grown under the same condition. All lists of differentially expressed sRNA genes are available in Appendix III.

For convenience of analysis of the SsrA/B regulon, differentially expressed genes in the  $\Delta ssrA$ ,  $\Delta ssrB$  and  $\Delta ssrAB$  mutant strains were combined as  $\Delta ssrAB$  in the following figures, however individual gene lists are available in Appendix III. Figure 3.7 B shows the strong overlap of differentially expressed genes from the three mutant strains and genes which are not shared between the three datasets are typically just under the 2.5-fold differential expression cut-off.

Of the 170 sRNA genes that are differentially regulated in any of the TF mutants, 44.1% (75 genes) are only differentially expressed in 1 regulatory mutant, 20% (34 genes) are differentially expressed in 2 regulatory mutants, 12.4% (21 genes) are differentially expressed in 3 regulatory mutants, 10% (17 genes) are differentially expressed in 4 regulatory mutants and 13.5% (23 genes) have regulatory input from 5 or more regulators (Figure 3.7 C).

Figure 3.7 D shows the number of sRNA genes which are differentially expressed in each regulatory mutant. The size of each bubble is indicative of the number of genes which are differentially regulated in an isogenic mutant of that regulator compared to wild-type grown under the same condition. The x-axis represents sRNA genes which are more highly expressed when the TF is absent (repressive regulatory input) and the y-axis represents genes which demonstrate a decrease in expression when the TF is absent

(activating regulatory input). The numbers of sRNA genes which are affected by the deletion of each TF are indicated in the legend to Figure 3.7 D.

There is large variation in the size of the sRNA regulons of each of the regulatory proteins investigated in this study. The absence of  $\sigma^{38}$  resulted in differential expression of 87 sRNA genes, 32 of which had been previously reported (Levi-Meyrueis et al., 2014, Frohlich et al., 2012, Silva et al., 2013). The large number of sRNA genes which are differentially expressed in the  $\Delta rpoS$  mutant, as well as the control of rpoS translation by 4 sRNAs (Majdalani et al., 1998, Majdalani et al., 2002, Mandin & Gottesman, 2010, Altuvia et al., 1997), highlights the importance of RpoS as a hub for post-transcriptional regulation.

The sRNA regulons of PhoB/R, RpoN, RpoE and HilD grown under SPI2-inducing conditions are much smaller by comparison. Only 1 sRNA, STnc1860 is differentially expressed in the absence of PhoB/R (>4-fold decrease in expression). STnc1860 is located 3' to the *phoU* gene which encodes a transcriptional regulator of the PstSCAB-PhoU high affinity phosphate transport system operon and is a target of the PhoB/R TCS (Wanner, 1996). STnc1860 appears to be co-transcribed with this operon from the *pstS* start site (Kröger et al., 2013) and could also play a role in phosphate assimilation.

The absence of RpoE and RpoN both result in the differential expression of 4 sRNA genes. As previously mentioned RpoE regulates expression of RybB and MicA (Papenfort et al., 2006) while RyeF appears to be a third sRNA in the *S.* Typhimurium RpoE regulon. Expression of MicC also decreases 2.5-fold in the *rpoE* mutant compared to wild-type grown to LSP. This is unsurprising as MicC also targets and silences *ompD* mRNA to help to maintain envelope homeostasis (Pfeiffer et al., 2009).

The 4 sRNA genes which are differentially expressed in the  $\Delta rpoN$  mutant are all more highly expressed in the absence of RpoN. As previously discussed, expression from the RpoN-dependent glmZ promoter increases in an rpoN mutant due to an overlapping RpoD-dependent promoter (Gopel et al., 2011) and this may be also be the case for these up-regulated sRNA genes. STnc2040 overlaps the 3' end of the narK gene which encodes a nitrite extrusion protein and is part of a group of anaerobically-induced nitrite extrusion proteins which are negatively controlled by RpoN in E. coli (Dong et al., 2011). STnc2040 is transcribed from its own transcriptional start site (Kröger et al., 2013) but may be co-regulated with narK. There is an up-regulation of fnrS in the  $\Delta rpoN$  mutant,

which also strengthens the link between nitrogen metabolism and anaerobic metabolism which has been demonstrated previously in *E. coli* (Birkmann *et al.*, 1987) and suggests that FnrS could provide a link between the FNR and RpoN regulons.

Five sRNA genes are differentially expressed in the  $\Delta hilD$  mutant grown under SPI2-inducing conditions. The low basal level of hilD expression under SPI2-inducing conditions in comparison to ESP (>20-fold decrease in hilD expression from ESP to inSPI2) is likely to account for the small numbers of HilD-regulated genes under SPI2-inducing conditions. Expression of GcvB, RyhB-2 and STnc870 increases approximately 3-fold in the  $\triangle hilD$  mutant. However, HilD has never been shown to directly repress any gene targets suggesting these regulatory interactions are likely to be indirect due to differential expression of other regulatory proteins. Two sRNA genes, STnc3750 and STnc3020, are approximately 3-fold reduced in the absence of HilD, under SPI2-inducing conditions. Both genes have multiple regulatory inputs (5 or more) and notably, expression of both genes is reduced in the majority of the SPI2-associated regulatory mutants grown under SPI2-inducing conditions. HilD activates SPI2 genes via regulation of ssrB during late stationary phase growth, but is not required for SsrB protein expression during growth in N-minimal medium (Bustamante et al., 2008). However, absence of HilD under SPI2-inducing conditions leads to an approximately 1.5-fold decrease in expression of the ssrA and ssrB genes. This decrease may not be detectable at the protein level by western blot (Bustamante et al., 2008), or may reflect the differences between N-minimal medium and PCN medium as inducers of SPI2 gene expression. This slight decrease in SPI2 regulator gene expression in the  $\Delta hilD$  mutant grown under SPI2-inducing conditions is likely to explain the reduction in expression of STnc3750 and STnc3020. However, STnc3020 is encoded on SPI1 and is antisense to prgI so direct binding and regulation by HilD at the STnc3020 promoter cannot be ruled out.

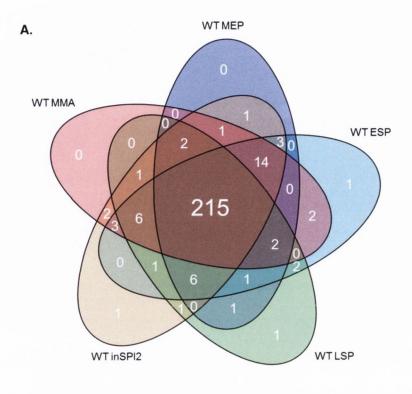
A general pattern of sRNA gene regulation emerges, in which we see larger numbers of differentially expressed sRNA genes in mutants of SPI2-associated regulators grown under SPI2-inducing conditions, compared to mutants of the invasion-associated regulators grown under SPI1-inducing conditions. Many of the sRNA genes which are differentially expressed in the SPI2-associated regulatory mutants are likely to be indirectly regulated by one or more SPI2-associated regulators due to the complexities of the SPI2 regulatory hierarchy (Fass & Groisman, 2009). We also see higher numbers of up-regulated sRNA genes than down-regulated sRNA genes in the mutant strains that

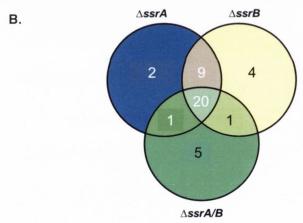
lack Fur and FNR, which is consistent with Fur primarily acting as a transcriptional repressor (Carpenter et al., 2009) and FNR being able to act either as a repressor or activator of transcription (Fink et al., 2007).

In general the effect of each TF, under investigation, on sRNA transcription is similar to the effect of each TF on transcription of protein-coding genes, in terms of the proportion of differentially expressed putative target genes (Figure 3.7 E). Again, there is a larger proportion of differentially expressed protein-coding genes in the absence of the SPI2-associated regulators, compared to the SPI1-associated regulators. The key difference in transcriptional regulation of putative sRNA gene targets and putative protein-coding gene targets occurs in the  $\sigma^{38}$  regulon.  $\sigma^{38}$  putatively regulates more sRNA genes than any of the other regulatory proteins. However the protein-coding-based regulon of  $\sigma^{38}$  is among the smaller regulons, containing only 213 putative target genes. As previously mentioned  $\sigma^{38}$  appears to be a hub for sRNA-mediated gene regulation and is clearly much more important for regulation of sRNA gene expression than for protein-coding gene expression, under the conditions of this study. The reason for the difference in regulon size may arise from the analysis of the  $\sigma^{38}$  regulon under LSP conditions, when many genes are considered not expressed and were filtered out of our analysis. The genes which are expressed and are regulated by  $\sigma^{38}$  at LSP are likely to mainly play a role in the general stress response (Hengge-Aronis, 2002). The importance of  $\sigma^{38}$  in regulation of sRNA expression at LSP reflects the well-documented role that sRNAs play in mediating cellular responses to stress (Vogel, 2009).

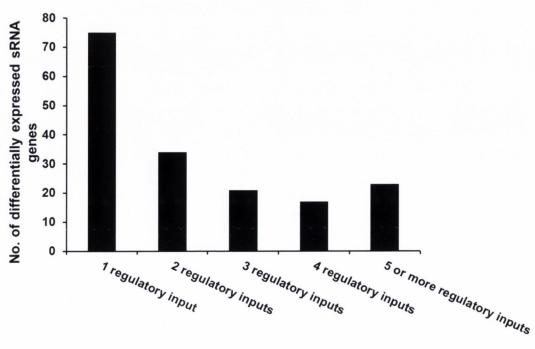
sRNA promoters are not qualitatively different from the promoters of protein-coding genes (Gottesman, 2005) and, with the exception of the  $\sigma^{38}$  regulon, the striking similarities of regulatory input on sRNA promoters with the promoters of protein coding genes highlight that the transcriptional regulation of sRNAs is likely to fit within established cellular regulatory networks. From our panel of important *S*. Typhimurium TFs, there does not appear to be a dedicated TF for transcriptional control of sRNAs, rather sRNAs are likely to have been co-opted into existing networks as the sRNA function was required.

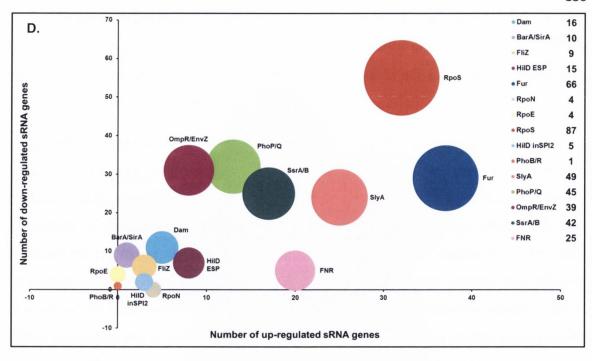
Figure 3.7 Global analysis of regulatory inputs on S. Typhimurium sRNA genes





C.





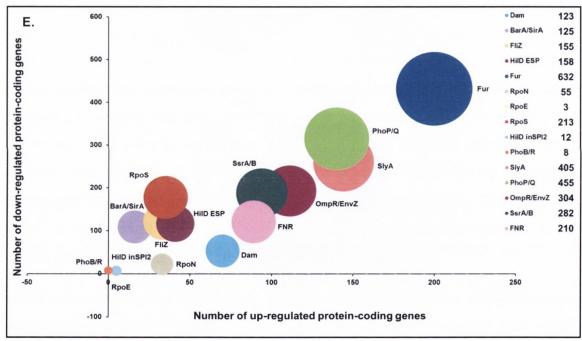


Figure 3.7 Global analysis of regulatory inputs on S. Typhimurium sRNA genes

**A.** Venn diagram demonstrating the overlap of sRNA genes which are expressed in wild-type under the 5 environmental conditions used in this study. **B.** Large overlap between the differentially expressed genes in  $\triangle ssrA$ ,  $\triangle ssrB$  and  $\triangle ssrA/B$  mutants so the three regulons were combined for convenience of analysis. **C.** Bar chart showing the numbers of differentially expressed sRNA genes in the context of the number of putative regulatory inputs to each gene. **D** and **E.** Bubble graph showing the numbers of differentially expressed sRNA genes (**D**) and protein-coding genes (**E**) in the isogenic mutant of each indicated regulator. The size of the bubble is proportional to the amount of differentially expressed genes in the isogenic mutant of the indicated regulator. The x-axis indicates the number of genes which are up-regulated in the mutant and the y-axis indicates the number of genes which are down-regulated in the mutant. The numbers of differentially expressed genes in each mutant strain are indicated in the figure key to the right of each graph.

#### 3.2.7 Global classification of differentially expressed sRNAs

The 280 known and candidate sRNAs in *S.* Typhimurium have previously been categorised into 5 classes based on their chromosomal location in relation to nearby coding genes (Kröger et al., 2013). The 5 classes are as follows: intergenic, 3' end-located, 5' end-located, antisense and intragenic. The antisense classification is based on the chromosomal location rather than mechanism of action of the sRNA, and sRNA genes were considered antisense if more than 50% of the transcript overlapped a gene on the opposite strand. Kröger *et al* found that 60% of sRNAs were encoded from intergenic regions, 20% were encoded 3' to a coding gene, 14% were antisense, 5% were encoded 5' to a coding gene and <1% were encoded intragenically. We used this classification to identify the chromosomal location of the 170 differentially expressed sRNAs to determine if certain mechanisms of transcriptional regulation are linked to particular chromosomal locations.

The differentially expressed sRNA genes show the same pattern of chromosomal location as the analysis by Kröger *et al* showed for all sRNA genes. 60.5% (103 genes) of differentially expressed sRNA genes are encoded from intergenic regions, 19.4% (33 genes) are 3'-end located, 14% (24 genes) are antisense-encoded, 5.3% (9 genes) are 5'-end located and 0.6% (1 gene) is intragenic (Figure 3.8 A). This pattern is mirrored for the small RNAs with a single regulatory input. The majority of differentially expressed sRNA genes, grouped according to their number of regulatory inputs, are encoded from intergenic regions. There are consistently high numbers of 3'-end located sRNAs in each group and the number of 3'-end located sRNA genes increases to represent approximately 1/3 of the sRNA genes with 4 or more regulatory inputs. sRNA genes located 5' to adjacent genes have no more than 3 regulatory inputs. There is an increase in the proportion of antisense transcripts to approximately 20% of the total proportion of sRNAs with 2 or 3 regulatory inputs. There are no antisense transcripts in the group of sRNAs with 4 regulatory inputs but over 17% of the sRNAs with 5 or more regulatory inputs are antisense-encoded (Figure 3.8 B).

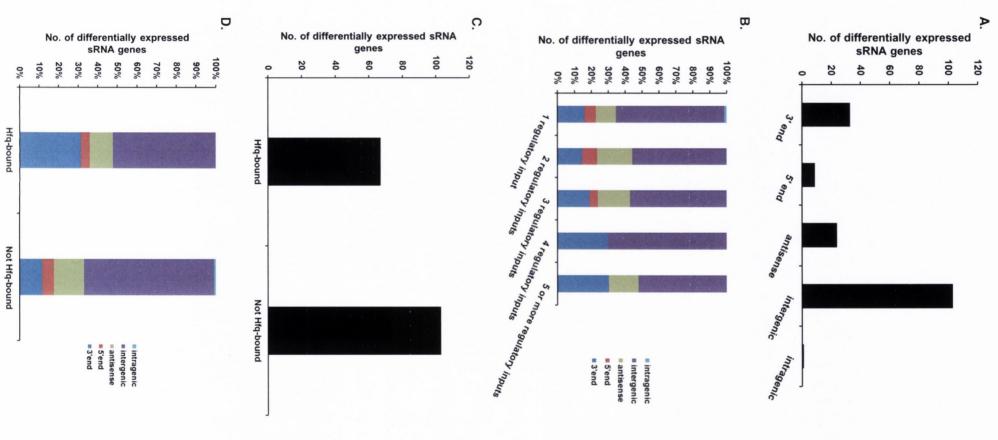
*Trans*-acting sRNAs generally require the RNA chaperone Hfq to aid in target binding (Vogel & Luisi, 2011). Kröger *et al* used data from Hfq co-immunoprecipitation studies (Sittka et al., 2008, Chao et al., 2012) to assess which of the known or candidate *S*. Typhimurium sRNAs were bound by Hfq and thus are likely to have a *trans*-acting

regulatory function. sRNAs were considered "Hfq-bound" if there was a Hfq-enrichment factor >5 over a mock immunoprecipitation in at least one environmental condition (Kröger et al., 2013). To determine if the differentially expressed sRNAs identified in this study are Hfq-bound, and are therefore more likely to adhere to the canonical models of *trans*-acting sRNA function, the differentially expressed sRNAs were categorised based on the analysis by Kröger *et al.* Almost 40% of the differentially expressed sRNAs were found to be strongly enriched for binding by Hfq while the remaining 60% were not Hfq-bound, at least under the conditions tested (Figure 3.8 C). This closely mirrors results from the Kröger *et al.* study which found that 41% of the 280 *S.* Typhimurium sRNAs were Hfq-bound.

Association with Hfq was also investigated in the context of the chromosomal location of differentially expressed sRNAs. 3'-end located sRNAs were more highly represented in the group of Hfq-bound differentially expressed sRNAs than in the non-bound group (31% of Hfq-bound differentially expressed sRNAs are 3'-end located compared to 11% of non-Hfq-bound differentially expressed sRNAs which are 3'-end located). This is unsurprising as the 3' untranslated region of genes has recently been shown to be an important reservoir for sRNA transcription (Chao et al., 2012) and the high representation of 3'-end located sRNAs in the group of Hfq-bound sRNAs indicate that these sRNAs are likely to function via a *trans*-acting mechanism. There is a slightly larger proportion of antisense-encoded sRNAs in the non-Hfq-bound group of differentially expressed sRNAs, suggesting that a number of these transcripts may function in *cis* and have perfect base-pair complementarity to their target mRNAs and do not require binding by Hfq for regulatory function.

The overall pattern of chromosomal location and requirement for binding by Hfq in the group of 170 differentially expressed sRNAs closely mirrors these patterns in the total number of *S*. Typhimurium sRNAs, indicating that chromosomal location and requirement for Hfq do not appear to influence the transcriptional regulation of sRNAs.

**Figure** expressed in at least one regulatory mutant 3.8 Global classification of sRNAs that were differentially



# Figure 3.8 Global classification of sRNAs that were differentially expressed in at least one regulatory mutant

**A.** Bar chart showing the number of differentially expressed sRNA genes in the context of their chromosomal location. **B.** Bar chart showing the percentage of differentially expressed sRNA genes grouped by the number of their regulatory inputs in the context of chromosomal location. **C.** Bar chart showing the number of differentially expressed sRNAs in the context of their binding to the RNA chaperone Hfq based on published enrichment scores from Hfq co-immunoprecipitation experiments (Chao et al., 2012, Kröger et al., 2013). **D.** Barchart showing the percentage of differentially expressed sRNAs in the context of their chromosomal location and whether they are bound or not bound by Hfq.

# 3.2.8 Motif analysis of differentially expressed sRNAs

Traditional methods of confirming direct regulation by a TF, such as electrophoretic mobility shift assays (EMSA) or DNase I footprinting, involve demonstrating that the TF can bind or protect a region of DNA containing the regulatory region of a target gene (Cai & Huang, 2012). These methods, often combined with nucleotide exchange experiments, have been used to define consensus binding motifs for TFs (Stapleton et al., 2002, Olekhnovich & Kadner, 2002, Walthers et al., 2007). Previously defined consensus binding motifs may then be used to predict other binding sites for the same TF, elsewhere on the chromosome, as previously described (section 2.10.2). Consensus binding sites for FNR and Fur have been identified in this way (Fink et al., 2007, Troxell et al., 2011a).

Chromatin immunoprecipitation (ChIP), followed by qPCR, microarray analysis or sequencing, has recently become a popular method of demonstrating physiologically relevant interactions between TFs and regions of DNA, as the assay is conducted *in vivo* (Cai & Huang, 2012). Another advantage of ChIP is that it can be used to study TF interactions with DNA on a global scale, unlike EMSAs and footprinting assays. Advances in ChIP and accompanying technologies have brought about a greater understanding of the DNA sequences required by bacterial TFs for binding DNA and gene regulation. Recent studies using ChIP-chip and ChIP-seq have also shown that some TFs have no preferred DNA sequence and can bind degenerate sequences or have preferred DNA topology rather than a sequence preference. HilD and OmpR are examples of this type of TF (Quinn et al., 2014, Petrone et al., 2014). However ChIP

using other TFs, such as SsrB, have successfully identified consensus motifs for binding (Tomljenovic-Berube et al., 2010). Well-defined consensus binding motifs, are useful tools to allow us to predict if putative regulatory interactions, such as those identified in our mutagenesis-based study, are direct or indirect interactions.

#### 3.2.8.1 Analysis of the Fur consensus binding motif and differentially expressed sRNAs

As previously discussed, Fur modulates expression of the sRNA RyhB. The direct Fur-mediated repression of RyhB allows Fur to indirectly positively regulate a number of genes necessary for bacterial survival and maintenance of iron homeostasis (Masse et al., 2005). This interaction is an important regulatory paradigm in understanding the interactions between TFs and sRNAs in cellular networks. We wished to investigate if Fur directly regulates the expression of other sRNA genes, which may play similar roles to RyhB within the Fur regulon. A position-specific scoring matrix (PSSM) was generated (section 2.10.2) using alignment of the homologous Salmonella sequences of from E. published Fur binding sites coli (available http://arep.med.harvard.edu/ecoli matrices/). Using software listed in section 2.10.3, these sequences were also used to generate a Fur consensus motif logo (Figure 3.9 A). The PSSM was used to scan for direct Fur binding, in the 100 bp upstream of the 37 sRNA genes which were up-regulated in the  $\Delta fur$  mutant, using pattern searching software and default settings, as previously described (section 2.10.2), as 100 bp was the maximum distance reported for Fur binding upstream of the published Fur-regulated genes that were used for matrix assembly. The published Fur binding sequences were scanned with the same PSSM to establish a minimum threshold weighted score, below which predicted motifs were considered to be false positives. However the lowest weighted score of a published Fur binding site was 7 and a number of poorly conserved motifs had a score of approximately 7. We, therefore, chose the more conservative threshold score of 10, and only motifs with a weighted score >10 were considered as putative Fur binding sites.

Fur binds as a dimer to a palindromic 19 bp "Fur box" which typically overlaps the promoter region of its target genes, although there is some uncertainty about the position and preferred DNA sequence for Fur binding to its target DNA (Escolar *et al.*, 1999, Baichoo & Helmann, 2002). The promoters of 6 up-regulated (repressed by Fur) sRNA

genes were found to contain putative Fur binding sites with a score above 10. The highest score (21.28) was obtained for a sequence within the RyhB-1 promoter that has previously been shown to be Fur-dependent in S. Typhimurium (Padalon-Brauch et al., 2008) (Figure 3.9 A). STnc4000 contains a putative Fur binding site overlapping the -10 site of the promoter region (score 10.04). This sRNA is encoded in an intergenic region and transcribed divergently from an iron-associated gene, bfd, which encodes a Fur-regulated bacterioferritin-associated ferredoxin protein (Figure 3.9 B). The second highest scoring putative Fur binding site (19.30) was obtained for a sequence overlapping the -10 site of the STnc3250 promoter. STnc3250 is also transcribed from an intergenic region, divergently from an iron-associated gene (fhuA) (Figure 3.9 C). fhuA is the first gene in an operon which encodes ferrichrome-iron associated proteins, which is under the control of Fur. The proximity of STnc4000 and STnc3250 to genes involved in iron homeostasis raises the possibility that Fur regulation of STnc4000 and STnc3250 may have evolved with the bfd gene and the fhuACDB operon, respectively, and that these sRNAs could play a role in maintenance of iron homeostasis. The Fur-repressed divergently oriented promoters of STnc3250 and STnc4000 may be new examples of the bidirectional promoters that have previously been reported in the Fur regulon, fepA-fes, fepD-ybdA and fepB-entCEBA. The fepA-fes promoter region contains overlapping Fur sites, allowing transcription of both genes to be modulated simultaneously by Fur (Escolar et al., 1998). In the case of the intergenic region upstream of STnc4000 and the divergently transcribed bfd gene, there are at least 2 Fur boxes present which do not overlap (Figure 3.9 B). However the proximity of the Fur boxes in the bidirectional promoter region suggests an efficient mechanism of coordinate regulation of gene expression by Fur. The STnc3250 and fhuA Fur boxes do overlap (Figure 3.9 C). The promoters of STnc3080, STnc3290, and STnc3140 also contain putative Fur binding sites, with respective scores of 12.39, 10.62 and 11.65 (Figure 3.9 A).

# 3.2.8.2 Analysis of the SsrB consensus binding motif and differentially expressed sRNAs

The identification of sRNAs which are directly regulated by SsrB could give us important clues about the role of these sRNAs in *S*. Typhimurium virulence, and could enhance our knowledge about the mechanisms employed by SsrB to integrate environmental signals

and mediate intracellular *S.* Typhimurium survival. ChIP-chip analysis of the SsrB regulon identified an 18bp palindromic sequence with internal 7-4-7 organisation (Tomljenovic-Berube et al., 2010). A PSSM was generated using the SsrB-bound sequences identified by Tomljenovic-Berube *et al.* These sequences were used to generate a consensus binding motif logo for SsrB (Figure 3.9 D). The PSSM was used to scan approximately 500bp upstream of the 21 sRNA genes which were down-regulated (2.5-fold or greater) in the absence of SsrB in this study. A score of 10 was used as the minimum threshold score for an SsrB consensus motif, based on scores of defined SsrB binding sites (Tomljenovic-Berube et al., 2010) scanned with the same PSSM.

The promoter regions of three differentially expressed sRNA genes contained one SsrB putative binding site, while one sRNA promoter contained two putative SsrB binding sites. Tomljenovic-Berube *et al* reported that SsrB could tolerate a quite degenerate palindrome sequence, particularly in the 4bp spacer region, but required at least one conserved heptamer in the correct orientation. An alignment of the predicted SsrB binding sites upstream of these 4 sRNAs (Figure 3.9 D) shows that the first heptamer in the palindrome is more highly conserved, while the spacer and second heptamer have more degenerate sequences.

A variable distance between the putative SsrB binding sites and the TSS of each sRNA was identified. The putative SsrB binding sites are located 204 bp and 410 bp upstream of STnc3210 and STnc3400, respectively, and similar distances were reported for other SsrB-dependent genes, such as *ssrB*, *srfN*, *pipB* and *steB* (Tomljenovic-Berube et al., 2010). The putative SsrB binding site was located 63 bp upstream of STnc3090, adjacent to the STnc3090 -35 site (Figure 3.9 F). There were 2 putative SsrB binding sites upstream of STnc3020, which is encoded antisense to *prgI* on SPI1. Site A was located 293 bp upstream of the STnc3020 TSS. Site B is located 53 bp upstream of the STnc3020 TSS, adjacent to the STnc3020 -35 site (Figure 3.9 E). The more proximal consensus site was also the highest scoring consensus site among the promoters of the differentially expressed sRNAs (18.36), raising the interesting possibility that SsrB binds within SPI1 and directly regulates a SPI1-encoded sRNA.

# 3.2.8.3 Analysis of the $\sigma^{38}$ consensus binding motif and differentially expressed sRNAs

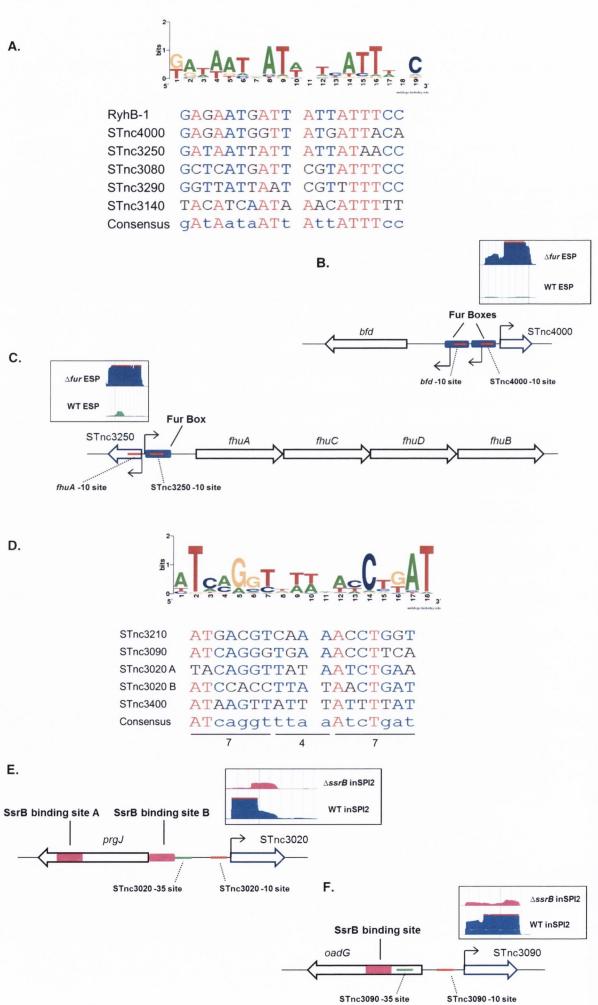
Absence of  $\sigma^{38}$  led to the largest numbers of differentially expressed sRNA genes. The  $\sigma^{38}$  and  $\sigma^{70}$  core consensus sequences are very similar, but deviations from the consensus are better tolerated by  $\sigma^{38}$  and are counter-selective for  $\sigma^{70}$ , allowing  $\sigma^{38}$  to specifically target and activate promoters under stressful conditions (Typas et al., 2007). promoter regions of the differentially expressed sRNAs in the  $\Delta rpoS$  mutant were aligned and analysed to determine if they contain hallmarks of  $\sigma^{38}$ -dependent promoters and are, therefore, likely to be part of the RpoS regulon. A C nt at position -13 relative to the TSS is highly conserved among  $\sigma^{38}$ -dependent promoters and is counter-selective for  $\sigma^{70}$ binding (Becker & Hengge-Aronis, 2001). A total of 32 genes (37%) of the sRNA genes which are differentially in the absence of  $\sigma^{38}$  contained a C at position -13 (Figure 3.9 G). Furthermore,  $\sigma^{38}$ -dependent promoters also typically carry a G or a T at position -14. A T nt in this position enhances selectivity for RpoS, but compromises the activity of the promoter, while a G in the -14 position does not affect promoter selectivity but does enhance promoter activity (Typas et al., 2007). In total, 44% of the differentially expressed sRNAs which have a C at position -13 had a G at position -14 and 12.5% contained a T at position -14 (Figure 3.9 G).

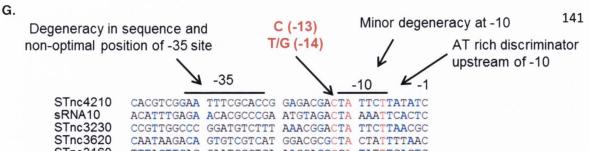
An AT-rich discriminator upstream of the -10 hexamer enhances  $\sigma^{38}$ -dependent transcription (Typas et al., 2007). The majority of sRNA promoters shown in Figure 3.9 G contain an AT-rich region immediately upstream of the -10 site. Unlike  $\sigma^{70}$ ,  $\sigma^{38}$  tolerates promoter sequence degeneracy.  $\sigma^{38}$ -dependent promoters are, therefore, typically less well-conserved than  $\sigma^{70}$ -dependent promoters. Minor sequence degeneracy in  $\sigma^{38}$ -dependent promoters is common at the -10 site, while the -35 site can vary widely both in sequence and position, leading to spacer lengths which would be sub-optimal for  $\sigma^{70}$  recognition and binding but can be utilised by  $\sigma^{38}$  (Typas et al., 2007). This sequence and location degeneracy is particularly evident in the alignment of promoters from differentially expressed sRNAs which shows virtually no conservation of their -35 hexamers (Figure 3.9 G).

The 32 sRNAs presented in Figure 3.9 G are likely to be  $\sigma^{38}$ -dependent as they contain strong indications of a preference for  $\sigma^{38}$  binding. However, promoters that have been experimentally shown to be activated by  $\sigma^{38}$  can contain all, or only some, of these

promoter features in different combinations, which affects overall promoter activity. I speculate that the other 55 sRNAs which were differentially expressed in the  $\Delta rpoS$  mutant, but do not contain a C at position -13, may contain other features that enhance  $\sigma^{38}$  binding. As an example, the *sraL* promoter does not contain a C at position -13 but ChIP has been used to demonstrate RpoS binding at the *sraL* promoter (Silva et al., 2013).

Motif analysis is an extremely useful tool for the development of hypotheses regarding direct transcriptional regulation of a gene by a TF. However, the use of EMSAs, DNase I footprinting or ChIP is required to validate these computational predictions as gene promoter regions contain various features and complexities that cannot be computationally predicted (Cai & Huang, 2012). The use of single nucleotide exchange mutagenesis can also be used to confirm the key nucleotides necessary for TF binding.





STnc3160 TTTACTTGAG GAATGGCTGA AAGCAGGCCA TATTTGACTG IstR-1,2 ATTTATAGAC AGAACGGAGC TTCTGATCCA TACTCAAATT STnc4140 CACGCTGACA TTAACACCTG TCTTGAACGA TAATCAATTT STnc3040 GTTCCACAGT ATGATGCGCG GTTTGCGCTA TCATCAATTT STnc1530 CGGCCATACA CTTTCATCAA CACCAGGCTA TCATGCTGTT STnc750 TTTTCGGGCG TTGTCTACTG TACCTGGCTA AACTGACTTC STnc1060 CTTTGTTGCA ATGGCAACGG TTCGCGGCTA TCCTCAACTT OmrA CGGACTGGCG ATGCAGACGC GCGAAGCCCA CAATCAAGAT STnc1410 AGCCCTTGCG CCAGACAGGC CACTCACCAA CAATCGCGAT STnc4010 GTTAGGCTTA ATACCCGTGG GATTAAGCTA CCCTTAAAGA SdsR GCAAACTGGA AAACCTGGCG TCGTCATCTA TTCTTAAAGG STnc3110 GTTTATCGCT AAAGCGTTAA GCGCAGTCTA TGATAACAGT IsrJ GGTGATTGCA AAACTGGATT AAAACGGCTA TATTTTATGT sRNA1 AACTGATGCA GTTCTGATTG ATCTGTACGC AGATATAAAC STnc2090 CGCCAGCGCA AAACACGAAG ATATGTACAC CGCTATTAAT STnc1330 CAAAGCAGCA GATGGCTTAA TACTTTCCTA TACTTTGTTT RygD CGCGGTTGAC GCTGCTTTTT TTTCGACCTA AAGTAAAGGA RyhB-2 ATGTCTTAAC GTTAGCGGTT TATTTGCCGT STnc1390 AAAATTGATT ACTTTGGCAG TGTTTGCCAT AATTACTCCT STnc1080 GAATTTGTAT ATTTCAGCGG TATTTGACAT ACTCATCTCC CTTAAGGTTC GCTTAATCTC TCGCGGGCAT ACTCTCCTCC STnc630 FnrS GTCAATAAAC CATCTACCTA TTCGGGGCAA TATCTCTCTC STnc2080 TTATCAGTAT GACTTTTTTC TGCGATGCTA TAGTTAATAT STnc4100 ACGCTAGTTT TCCAATATAC ATATATGCAA TATTTGCAGC Isrl TCTAACGCCA ACCTGCAATC CCAATAGCTA AACTCCTCTT STnc1560 CCATCGCCCT GACGGCGATC CAGGCTTCTC AGCAGCAGCA STnc290 ATGTTGACTG GAGATATCTG CCAACAGCAT AACTTTTACG SroG TTCTCATCAC GGCGCGCCGT GACGGAACGC GGAGAGCAGG Consensus ctttcttgca at..c..ct. ta.t.ggCtA tacTtaaatt

Figure 3.9 Motif analyses of differentially expressed sRNA promoters

A. Motif consensus logo for Fur recognition site. Multiple alignment of high-scoring Fur consensus sites in the promoter region of sRNA genes which were up-regulated in the  $\Delta fur$  mutant. **B.** Genetic organisation of STnc4000 and bfd promoter regions showing predicted Fur boxes (blue boxes) overlapping the -10 sites of each gene. White arrows with black and blue outlines show protein-coding and sRNA genes, respectively. Bent black arrows denote TSS. All arrows indicate the direction of transcription. Red bar indicates -10 hexamer. Inset figure shows jbrowse visualisation of sequencing reads for STnc4000 from wild-type and  $\Delta fur$  mutant at ESP. The height of the track is directly proportional to gene expression. C. Genetic organisation of STnc3250 and fhuA promoter regions showing predicted Fur boxes. Symbols and inset image as in (B). D. Motif consensus logo for SsrB recognition site. Multiple alignment of high-scoring consensus sites in the promoter region of sRNA genes which were down-regulated in the ΔssrB mutant. E. Genetic organisation of STnc3020 promoter region and predicted SsrB binding sites (pink boxes). Symbols as in (B) and green bar indicates -35 hexamer. Inset figure shows jbrowse visualisation of sequencing reads for STnc3020 from wild-type and ΔssrB mutant at inSPI2. F. Genetic organisation of STnc3090 promoter and predicted SsrB binding site. Symbols and inset image as in (E). Binding sites predicted, as previously described (section 2.10.2). G. Multiple alignment of promoter regions of sRNA genes which were differentially expressed in the  $\Delta rpoS$  mutant and showed the strongest consensus motif for  $\sigma^{38}$  binding. Some promoter features that favour recognition by  $\sigma^{38}$  are indicated above the alignment. For each alignment: Red >75% nucleotide sequence identity, Blue >35%<75% nucleotide sequence identity, Black <35% nucleotide sequence identity between each aligned promoter sequence.

### 3.2.9 Transcriptional network analysis

The high-throughput analysis of differential sRNA gene expression in 17 TF mutants in this study has made it possible to generate a transcriptional regulatory network. Global regulatory networks are useful tools to provide insight into the systems, components and regulatory interactions that may not be apparent by investigating individual components of the network in isolation (McDermott et al., 2009). The transcriptional regulatory network, based on differential expression of 170 sRNAs which show a 2.5-fold or greater change in expression in one or more regulatory mutant, was generated using Cytoscape software (www.cytoscape.org) (Figure 3.10). Hubs are highly connected proteins within the regulatory network (McDermott et al., 2009). In Figure 3.10 each TF is a hub and TFs are represented as coloured hexagons. sRNAs are represented as pale blue circles. Regulatory interactions are indicated by red arrows or blue T-shaped lines. The red arrows indicate putative activation of sRNA transcription by a regulator (expression of the sRNA is reduced in a mutant strain that lacks a particular regulator). The blue T-shaped lines indicate putative repression of sRNA transcription by a regulator (expression of the sRNA is increased in a mutant strain that lacks a particular regulator). These lines that represent regulatory interactions are known as edges. The network is organised using an edge-weighted force-directed layout, which imposes the weight of the fold-change in expression of each regulatory interaction upon each edge. Therefore the length of each edge indicates of the strength of the regulatory interaction. sRNAs that showed a higher fold-increase or decrease in expression in a particular mutant strain that lacked a TF are located closest to that TF node, while an increasing distance of the sRNA node from the TF node indicates a smaller change in expression of that sRNA in the mutant strain lacking the particular regulator.

The PhoB/R TCS and the PhoB/R-dependent sRNA, STnc1860, are not connected to the remainder of the regulatory network. As previously discussed, STnc1860 is encoded 3' to a PhoB/R-dependent operon and the lack of a significant change in STnc1860 expression in any other regulatory mutant emphasises the importance of the PhoB/R TCS for expression of this sRNA. The fact that deletion of the *phoB/R* genes does not result in significant differential expression of any other sRNA, even under the phosphate limiting conditions of SPI2-inducing PCN, suggests that sRNAs do not play a central role in phosphate homeostasis.

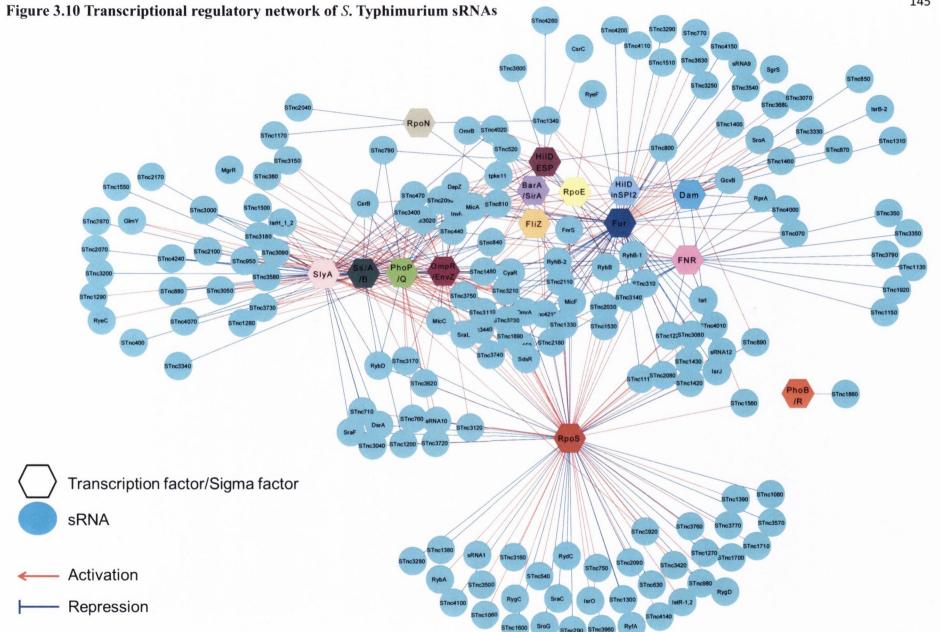
The remainder of the regulatory network highlights the complexity of the transcriptional regulation of sRNAs and some regulatory patterns clearly emerge. The SPI2-associated regulators cluster together in the most complex area of the network. This complexity reflects the amount of shared sRNA gene targets among this group although some of these interactions are likely to be indirect. A number of sRNA genes which are down-regulated in the SPI2-associated mutants are up-regulated in the mutant that lacks FNR. As previously mentioned, SPI2-associated regulators and effector genes are de-repressed in the absence of FNR under the conditions tested in this study. The similar pattern of de-repression of SPI2-associated regulator-dependent sRNAs in the strain that lacks FNR may either be due to indirect effects resulting from up-regulated *ssrB* expression or these sRNAs may be controlled by the same mechanism as FNR uses to repress SPI2 gene expression. This will be the subject of further experimentation.

The network shows a large number of shared interactions between RpoS and the SPI2-associated regulators. A preference for  $\sigma^{38}$ -dependent promoters among sRNAs which are dependent on SPI2-associated regulatory mutants may reflect the starvation and stress-inducing conditions which activate SPI2 gene expression. The absence of the RpoS and Fur proteins result in the largest number of differentially expressed sRNA genes, and both RpoS and Fur are putatively involved in the regulation of a number of sRNA genes that have no other regulatory inputs, highlighting the importance of these regulators in control of sRNA gene expression.

The SPI1-associated TFs also form a regulatory cluster. However, the SPI1-associated regulatory cluster is more diffuse than that of the SPI2-associated regulators, suggesting that there is not as much co-regulation and cross-talk between the regulons of the SPI1-associated TFs.

The previously identified sRNAs with multiple regulatory inputs connect multiple hubs, and we can speculate that these sRNAs play a physiological role in connecting these regulators to integrate multiple regulatory signals to generate a co-ordinated genetic output *in vivo*. The sRNAs with fewer regulatory inputs are more in isolated locations of the regulatory network and it may be speculated that these sRNAs play more specific roles within particular regulators. Algorithms, such as context likelihood of relatedness (CLR), have been developed to increase the confidence and precision of inferring information from transcriptional regulatory networks, such as Figure 3.10. CLR scores mutual

expression values for two genes, such as a TF and a target gene, from diverse datasets to calculate the statistical likelihood of direct regulatory interactions (Faith *et al.*, 2007). In the future, studies containing high-throughput data regarding the expression of *S*. Typhimurium sRNA genes under a variety of environmental in vitro conditions (Kröger et al., 2013, Chao et al., 2012, Sittka et al., 2008) and further studies, perhaps under *in vivo* conditions, can be combined with the transcriptomic data from this study to improve the precision of inferring physiologically relevant regulatory interactions.



# Figure 3.10 Transcriptional regulatory network of S. Typhimurium sRNAs

Regulatory network was generated using software from <a href="www.cytoscape.org">www.cytoscape.org</a>. Regulatory interactions are based on mutant RNA-seq data. sRNA genes that are down-regulated in a mutant strain lacking a certain TF or  $\sigma$ -factor (regulatory protein) are putatively activated by that regulatory protein, and the interaction is represented using a red arrow from the source node (regulatory protein) to the target node (sRNA). sRNA genes that are up-regulated in a mutant lacking a certain regulatory protein are putatively repressed by that regulatory protein, and the interaction is represented using a blue T-shaped line from the source to the target node. The length of each line is representative of the strength of the putative interaction i.e. nodes representing sRNAs which show higher fold-increase or decrease in a mutant strain are located closer to that source node. Source nodes are hexagonal in shape and have a variety of colours. Target nodes are circular and pale blue in colour. All regulatory interactions between TFs and sRNA genes are detailed in Appendix III.

# 3.2.10 Identification of SPI1 and SPI2-like sRNAs

The "guilt by association" hypothesis posits that groups of genes which perform similar functions should be co-expressed and/or co-regulated. Using this hypothesis, it should be possible to define patterns of expression of functional groups of genes, and to identify new genes which share the same function, based on the pattern of expression of these genes (Quackenbush, 2003). In order to find sRNA genes which may play important roles in S. Typhimurium virulence we adopted the "guilt by association" hypothesis to find sRNAs which show similar patterns of regulation to SPI1 and SPI2 genes. Genespring software (Agilent Technologies) provides a "cluster analysis" programme, which is used to group genes together based on their pattern of expression in a number of conditions/mutant strains. Cluster analysis of all 280 S. Typhimurium sRNAs based on their regulatory patterns in 15 TF mutants, across 5 environmental conditions, shows that the patterns of differential sRNA gene expression mirror the general pattern of differential protein-coding gene expression (Figure 3.11, compare panels A and B). SPI1, SPI2 and flagellar genes have distinct patterns of expression (Kröger et al., 2013) and regulation, as previously discussed in chapter 1 (Ellermeier & Slauch, 2007, Fass & Groisman, 2009, Chilcott & Hughes, 2000). The SPI1, SPI2 and flagellar gene clusters are highlighted in Figure 3.11 A. The Pearson correlation coefficient was used to identify groups of sRNAs which show a similar expression profile over the panel of TF mutants. Correlative analysis was used to measure similarity between these groups of sRNAs to archetypical SPI1 (prgH) or SPI2 (ssaG) genes across the panel of regulatory mutants. The prgH and ssaG genes are required for the function of either SPI1 or SPI2, and report

the output of the regulatory hierarchy that controls expression of each TTSS (Kröger et al., 2013). sRNA genes that shared a correlation coefficient >0.7 with *prgH* or *ssaG* were considered to have "SPI1-like" or "SPI2-like" transcriptional signatures (section 2.10.1).

#### 3.2.10.1 SPI1-like sRNAs

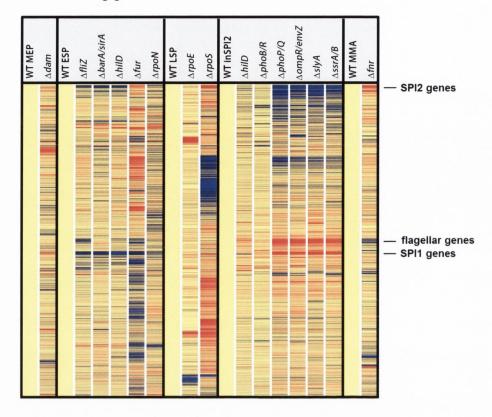
Four sRNAs show a SPI1-like pattern of expression. InvR and DapZ have correlation coefficients of 0.76 and 0.7 respectively. Expression of both of these sRNAs is dependent on the primary SPI1 regulator HilD (Pfeiffer et al., 2007, Chao et al., 2012). Expression of STnc520 was highly correlated with expression of prgH (Pearson correlation coefficient: 0.87) and STnc520 is also HilD-dependent (Figure 3.6 A). InvR, DapZ and STnc520 are grouped with SPI1 genes and genes which are co-regulated with SPI1 (siiABCDEF), based on their pattern of regulation (Figure 3.11 C). Expression of a fourth sRNA, STnc1340, also correlates with prgH (correlation coefficient: 0.78). Similarly to expression of InvR, DapZ and STnc520 expression of STnc1340 decreases in a  $\Delta hilD$  mutant and increases in a  $\Delta slyA$  mutant. Of these 4 SPI1-like sRNAs, only InvR is encoded on SPI1. The correlation coefficient scores of each of the SPI1-like sRNAs are summarised in Table 3.2.

#### 3.2.10.2 SPI2-like sRNAs

Ten sRNAs show a SPI2-like pattern of expression. The sRNA genes STnc3180, STnc3430 and STnc380 correlate well with *ssaG* (correlation coefficients >0.7). The sRNA genes STnc3090, STnc3730 and STnc440 share correlation coefficients >0.8 with *ssaG*. The sRNA genes *isrH\_1\_2*, STnc1480, STnc3020 and STnc470 have the highest degree of correlation with *ssaG* (correlation coefficients >0.9). Figure 3.11 D shows a region of the SPI2 gene cluster containing SPI2 effectors, SPI2-associated genes and the sRNAs STnc3020, STnc1480 and STnc440. Northern blot analysis has confirmed that STnc1480 has multiple regulatory inputs from SPI2-associated regulatory proteins (Figure 3.6 E). None of the SPI2-like sRNAs are encoded on SPI2 but the promoter regions of STnc3020 and STnc3090 both contain putative SsrB binding sites (Figure 3.9). The correlation coefficient scores of each of the SPI2-like sRNAs are summarised in Table 3.2.

Figure 3.11 Cluster analyses of all protein-coding and sRNA genes based on their patterns of relative expression in 15 regulatory mutants

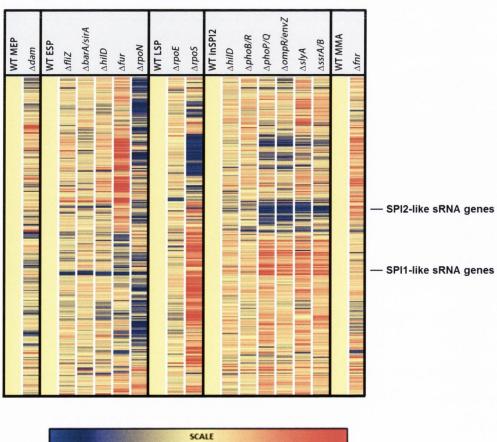
# A. Protein-coding genes



# B. sRNA genes

Down-

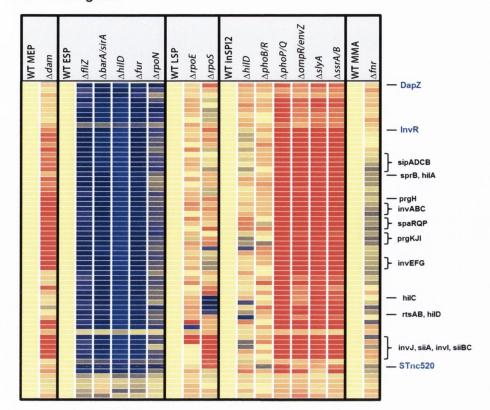
Regulated >6-fold



No

Change

Up-Regulated >6-fold



#### D. SPI2-like genes

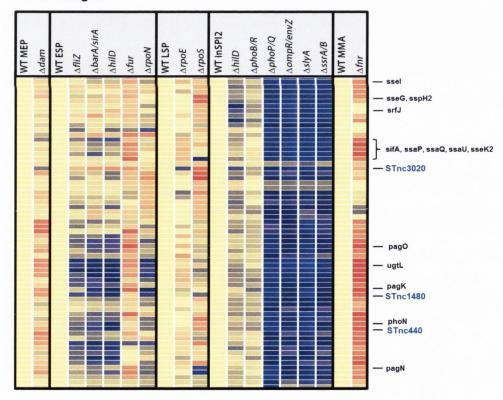


Figure 3.11 Cluster analyses of all protein-coding and sRNA genes based on their patterns of relative expression in 15 regulatory mutants

**A.** Heatmap showing relative expression of protein coding genes in mutants compared to wild-type, grown under the same condition. **B.** Heatmap showing relative expression of sRNA genes in mutants compared to wild-type, grown under the same condition. **C.** SPI1 genes and genes which are regulated similarly to SPI1 genes (SPI1-like). **D.** SPI2 genes and genes which are regulated similarly to SPI2 genes. sRNA gene names are labelled in blue. Each row represents expression of a single gene. In all heatmaps blue indicates a decrease in expression compared to wild-type (yellow), and red indicates an increase in expression compared to wild-type (see colour bar).

SPI1-like sRNA	Upstream Gene	Downstream Gene	Genomic context <sup>a</sup>	Pearson correlation coefficient <sup>b</sup> with prgH	TraDIS <sup>c</sup> attenuation
DapZ	dapB	STM0065	>>>	0.70	No
InvR	invH	STM2901	>>>	0.76	Yes
STnc1340	bcr	yejG	< < <	0.78	No
STnc520	pliC	pagC	> < <	0.87	Yes

SPI2-like sRNA	Upstream Gene	Downstream Gene	Genomic context	Pearson correlation coefficient with ssaG	TraDIS attenuation
STnc3430	STM1254	STM1253	<<<	0.72	No
STnc3180	ybdO	dsbG	<><	0.78	No
STnc380	lctP	sadA	><>	0.78	No
STnc3730	<i>yfcC</i>	pta	><>	0.84	No
STnc440	STM4310	STM4312	>><	0.85	No
STnc3090	oadG	STM0057	<><	0.89	No
STnc1480	yeaJ	yeaH	<><	0.91	Yes
STnc470	STM0082	STM0081	><<	0.92	No
STnc3020	prgJ	prgH	as <sup>d</sup> to <i>prgI</i>	0.93	Yes
IsrH_1_2	sseL	glpC	><>	0.99	Yes

<sup>&</sup>lt;sup>a</sup> Genomic context of sRNA relative to the flanking protein-coding gene. The middle arrow represents sRNA gene, direction of arrows represents direction of transcription

<sup>&</sup>lt;sup>b</sup> Pearson correlation coefficient calculated based on expression of sRNA in wild-type 4/74 and 17 regulatory mutants grown under 5 environmental conditions

<sup>&</sup>lt;sup>c</sup> Transposon insertions located within sRNA gene leads to significant attenuation within the chicken, pig or calf infection models (Chaudhuri *et al.*, 2013).

d as: antisense

#### 3.2.11 Conservation of SPI-like sRNAs

Salmonella diverged from its last common ancestor with E. coli over 100 million years ago, and evolved as a pathogen through acquisition of the SPI1 TTSS, however the ancestral (core) Salmonella chromosome shares a lot of similarities with the E. coli chromosome (Sabbagh et al., 2010). Salmonella enterica acquired the SPI2 TTSS following the evolutionary divergence from S. bongori (Porwollik & McClelland, 2003). We wanted to investigate if the SPI-like sRNAs were acquired at the same as the cognate SPI gene clusters, or during a separate lateral gene transfer event. Another possibility is that the SPI-like sRNAs are ancestral genes that were co-opted by the SPI-associated regulatory proteins. Investigating the conservation of the SPI-like sRNAs may provide information about the roles played by these sRNAs in S. Typhimurium virulence. As shown in Figure 3.2, the majority of the SPI1- and SPI2-like sRNAs are not encoded on SPI1 or SPI2. However the S. Typhimurium chromosome is mosaic in nature and contains a number of horizontally acquired regions of DNA or islands in addition to the main SPIs (McClelland et al., 2001, Sabbagh et al., 2010). The distribution of SPI1- and SPI2-like sRNA genes among horizontally acquired regions of DNA or the ancestral Salmonella chromosome was investigated, as well as their conservation within enteric bacterial species to understand how the regulation of these sRNAs may have evolved.

# 3.2.11.1 Conservation of the SPI1-like sRNAs

InvR is the only SPI1-like sRNA which is encoded on the SPI1 island. InvR is encoded at the end of SPI1 but appears to have been acquired with SPI1 and is under the control of the primary SPI1 regulator, HilD (Pfeiffer et al., 2007). DapZ is also positively regulated by HilD and is located 3' to the core ancestral *Salmonella* gene *dapB*. STnc520 is encoded on SPI11. STnc1340 is encoded intergenically to two core ancestral *Salmonella* genes (McClelland et al., 2001). The genomic location of each of the SPI1-like sRNAs is summarised in Table 3.2.

Overall, the group of SPI1-like sRNAs are highly conserved (>90% sequence identity) in serovars of *Salmonella enterica* and are partially conserved (67%-79% sequence identity) in the more distantly related *Salmonella bongori* (Figure 3.12 A). With the exception of STnc1340, the SPI1-like sRNAs are poorly conserved outside of the *Salmonella* genus.

The SPI1-encoded InvR, in particular, is only present in Salmonella species and is not found in any other enteric bacteria. Interestingly, the sequence of the 3'-derived dapZ gene is not highly conserved outside of Salmonella while the dapB sequence is well-conserved in γ-proteobacteria (Chao et al., 2012). The chromosome of the pathogenic E. coli O157:H7 contains a sequence with approximately 44% sequence identity to the S. Typhimurium STnc520 sRNA gene, while the chromosome of the non-pathogenic E. coli K-12 does not contain any region of homology to STnc520. STnc1340 has the highest conservation in Salmonella bongori of the four SPI1-like sRNAs. Both E. coli strains, Shigella flexneri and Citrobacter koseri contain sequences with approximately 80% sequence identity to the S. Typhimurium STnc1340 gene, while there is low sequence identity (< 45%) with the S. Typhimurium STnc1340 in the Klebsiella pneumoniae or Yersinia pseudotuberculosis chromosomes. The location of STnc1340 in the intergenic region between two core Salmonella ancestral genes and its high conservation in Salmonella bongori and E. coli, indicates that STnc1340 was not a recently acquired feature and its co-regulation with SPI1 genes may reflect co-option by SPI1-associated regulators to aid in Salmonella survival or invasion of the host gut.

# 3.2.11.2 Conservation of the SPI2-like sRNAs

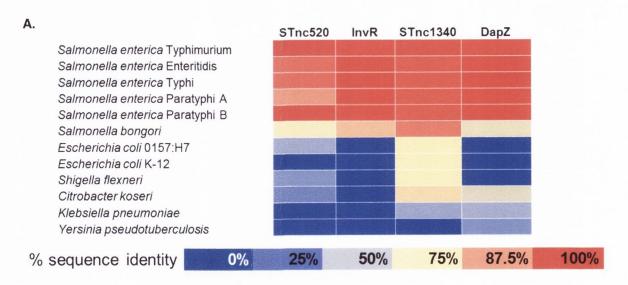
None of the SPI2-like sRNAs are encoded on the SPI2 island. As previously mentioned, STnc3020 is encoded on SPI1, antisense to *prgI* and is only expressed under macrophage-mimicking conditions, anti-correlate to *prgI*. *isrH2* is encoded within *isrH1* (Padalon-Brauch et al., 2008) so the TPM values of each gene cannot be separated and must be calculated as a single gene. IsrH\_1\_2 overlaps the 5' end of the *sseL* SPI2 effector gene and both genes were found to be inversely expressed in exponential and stationary phases in LB medium (Padalon-Brauch et al., 2008), however under the conditions tested here, both genes require OmpR and SsrA/B for optimal expression. STnc440, STnc470 and STnc3430 are encoded on small *Salmonella*-specific islands (McClelland et al., 2001). STnc3090 is encoded adjacent to a *Salmonella*-specific gene *oadG*. STnc3730 and STnc1480 are encoded adjacent to genes which are part of the core chromosome that have homologues in enteric bacteria outside of *Salmonella* (McClelland et al., 2001). STnc380 and STnc3180 are encoded in intergenic regions between *Salmonella*-specific genes *sadA* and *ybdO* respectively, and genes which are part of the

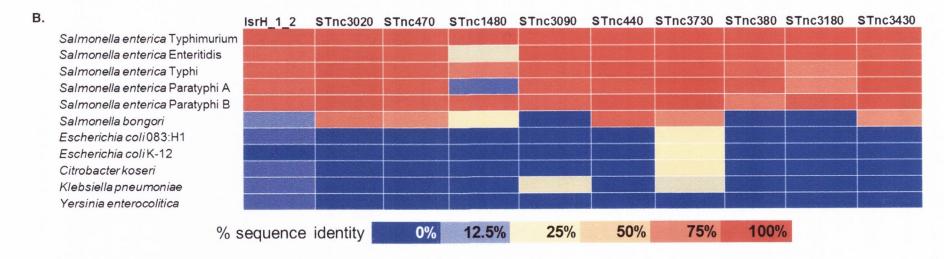
ancestral *Salmonella* chromosome (*lctP* and *dsbG* respectively) (McClelland et al., 2001). The genomic location of each of the SPI2-like sRNAs is summarised in Table 3.2.

The SPI2-like sRNAs are generally highly conserved (approximately 90%) within the Salmonella enterica serovars but the majority of SPI2-like sRNAs are less well-conserved in Salmonella bongori that the SPI1-like sRNAs are (Figure 3.12 B). STnc3730 is an exception among this group as STnc3730 has approximately 20% sequence identity with sequences in pathogenic and non-pathogenic E. coli as well as Citrobacter koseri and Klebsiella pneumonia. The location of STnc3730 adjacent to a core ancestral gene indicates that STnc3730 was not recently acquired. STnc3020 shows a similar pattern of conservation to InvR, which is unsurprising as both are encoded on SPI1 and are likely to have co-evolved despite their differential regulation by SPI2 or SPI1 regulatory proteins, respectively. STnc1480 is exceptional among the group of SPI2-like sRNAs as it is not STnc1480 has as highly conserved among all Salmonella enterica serovars. approximately 20% sequence identity with a region on the S. Enteritidis chromosome and the STnc1480 sequence is not present in S. Paratyphi A. Regions with sequence identity to IsrH\_1\_2, STnc3090, STnc380 and STnc3180 are not found in Salmonella bongori, suggesting that these sRNA genes were either lost from Salmonella bongori or that they were acquired by Salmonella enterica after the evolutionary divergence of the two species. The SPI2-like pattern of regulation of IsrH\_1\_2, STnc3090, STnc380 and STnc3180 favours the latter scenario may be correct.

We have identified 14 sRNA genes which are regulated similarly to archetypical SPI1 and SPI2 genes. The majority of these SPI-like sRNAs are specific to the *Salmonella* genus, and the SPI2-like sRNAs are, in general, restricted to the *Salmonella enterica* species, with some exceptions as discussed. Additionally, a number of the *Salmonella*-specific SPI-like sRNAs are located on sections of the chromosome that bear hallmarks of horizontal acquisition (McClelland et al., 2001). It is tempting to speculate that these sRNAs, which show the same patterns of regulation as important virulence genes, and that are generally not found outside of the *Salmonella* genus, play a role in the survival of *Salmonella* in specific niches or that the SPI-like sRNAs contribute directly to virulence through regulatory interactions with virulence factors. Further characterisation of the SPI-like sRNAs will be required to investigate these possibilities.

Figure 3.12 Conservation analyses of SPI1-like and SPI2-like sRNAs





**A.** Heatmap shows conservation of 4 SPI1-like sRNAs in the genomes of 12 enteric bacteria. **B.** Heatmap shows conservation of 10 SPI2-like sRNAs in the genomes of 11 enteric bacteria. Sequence identity was determined using BLAST software (section 2.10.3). The % conservation is compared to the ST4/74 sRNA sequence (100%), and is defined over the full length of the sRNA gene. Blue indicates no sequence identity; red indicates 100% sequence identity (see colour bars).

# 3.2.12 Investigation of the virulence-associated phenotype of SPI1- and SPI2-like sRNAs

To investigate if the SPI1-like and SPI2-like sRNAs play important roles during *S*. Typhimurium infection, published data from a transposon-directed insertion site sequencing (TraDIS) study was interrogated. High-throughput sequencing of insertion sites of pools of *S*. Typhimurium 4/74 transposon mutants, following oral infections of chicken, pigs and calves, in comparison to an input inoculum (Chaudhuri et al., 2013). The ratio of input to output reads was used to calculate the fitness score of each mutant. Transposon insertions that result in significant attenuation of the mutant strain were located in two SPI1-like sRNAs (InvR and STnc520) and in three SPI2-like sRNAs (STnc1480, STnc3020 and IsrH\_1\_2) (Table 3.2), consistent with these five sRNAs playing an important role during *S*. Typhimurium infection.

#### 3.2.13 Analysis of sRNA regulon of Hfq

As previously mentioned, the transcriptomic analysis of the *Salmonella* Hfq regulon was previously investigated by DNA microarray (Sittka et al., 2008). At that time, only 64 sRNAs had been identified in *S.* Typhimurium and these were not present on the microarray. Co-immunoprecipitation analysis by Sittka *et al* demonstrated that approximately half of the known or candidate sRNAs were bound by Hfq at ESP. This study was expanded recently when Chao *et al* did Hfq co-IP from a number of environmental conditions with increased numbers of sRNAs (Chao et al., 2012). In addition to acting as an RNA chaperone that facilitates and stabilises binding between sRNA and target mRNA molecules, Hfq also stabilises sRNAs through protection from ribonucleases prior to target recognition (Vogel & Luisi, 2011). However neither of the co-immunoprecipitation-based studies addressed how the absence of the Hfq protein affects the steady state levels of the Hfq-bound and non-bound sRNAs, which we

addressed in this study. However, the interaction between Hfq and other proteins or protein complexes, as well as the interaction of Hfq with many mRNAs, can lead to apparent effects on gene expression that cannot be directly attributed to Hfq (Wagner, 2013). Therefore, the previously published Hfq co-immunoprecipitation data (Chao et al., 2012) were combined with the transcriptomic analysis data from this study, to determine which differentially expressed sRNAs are also physically associated Hfq, and thus are likely to be canonical *trans*-acting sRNAs.

As previously described (section 3.2.7), a 5-fold enrichment factor of Hfq immunoprecipitation over a mock immunoprecipitation (Kröger et al., 2013) was used to determine if sRNAs were associated with Hfq at ESP, as this was the condition of growth for the wild-type and  $\Delta hfq$  strains in this study. Figure 3.13 A shows that 27% (75 genes) of all S. Typhimurium sRNA genes are differentially expressed in the absence of Hfq in this study (2.5-fold or greater change in transcript level). 80% of genes which are differentially expressed in the absence of Hfq were bound by Hfq in at least one of the conditions used by Chao *et al*, and over 61% (46 genes) are bound by Hfq at ESP.

Fourteen sRNAs are differentially expressed in the  $\Delta hfq$  mutant at ESP and are associated with Hfq under one or more of the conditions used by Chao *et al*, but are not Hfq-bound at ESP. A number of these sRNAs, such as ArcZ, DsrA, RybB and SgrS, are very highly enriched for Hfq binding under some conditions and the enrichment factors for Hfq binding to these sRNAs at ESP is generally just below the 5-fold enrichment threshold. The active cycling model of Hfq binding to sRNAs suggests that Hfq-bound sRNAs are replaced by binding of competitor sRNAs to Hfq, as a result of changing concentrations of RNAs (Wagner, 2013). In addition, Hfq protein is limiting in the bacterial cell, resulting in competition for binding among sRNAs (Moon & Gottesman, 2011). For this reason, altered levels of certain sRNAs throughout growth could affect the levels of, or the Hfq-association of other sRNAs which have a lower binding affinity for Hfq. Exchange of sRNAs on the available pool of Hfq occurs throughout growth (Chao et al., 2012) and upon over-expression of a Hfq-bound sRNA (Papenfort *et al.*, 2009). These features of Hfq binding to sRNAs are likely to account for the low enrichment of Hfq binding to ArcZ, DsrA, RybB and SgrS at ESP.

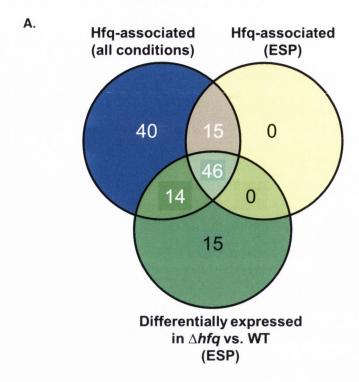
Fifteen sRNAs were bound by Hfq at ESP but were not differentially expressed in the  $\Delta hfq$  mutant at ESP. Two of these sRNAs, AmgR and STnc2100, were not expressed at

ESP or in the  $\Delta hfq$  mutant and so were excluded from our analysis. The remaining 13 Hfq-bound, non-differentially expressed sRNAs may require Hfq binding for their activity and to aid binding to their target mRNAs but do not require Hfq for stability. It is not known why Hfq does not always stabilise Hfq-bound sRNAs. GcvB is a relatively unstable sRNA and its stability is unaffected in the absence of Hfq, but GcvB requires Hfq for activity (Moon & Gottesman, 2011, Sharma et al., 2011).

Forty sRNAs are bound by Hfq under some conditions but not at ESP and unsurprisingly the levels of these genes are not affected in the absence of Hfq at ESP. These 40 sRNAs are unlikely to function at ESP and thus, are unlikely to compete for the limiting amounts of available Hfq under this condition. There are a further 15 sRNAs which are differentially expressed in the  $\Delta hfq$  mutant but are not associated with Hfq in any of the conditions used by Chao *et al*. The differential expression of a number of these sRNAs may be due to indirect effects, as the absence of Hfq has pleiotropic effects on  $\sigma$ -factors and TFs (Sittka et al., 2008), so the differential expression of some sRNAs in the hfq mutant may be as a result of altered transcription rather than an altered rate of turnover.

Indirect regulation by Hfq is particularly evident in the group of sRNAs whose levels increase in the absence of Hfq. Approximately 20% of differentially expressed sRNAs in the  $\Delta hfq$  mutant are up-regulated in the absence of Hfq (Figure 3.13 B). Among this group are the RpoE-dependent sRNAs MicA and RyeF. RpoE and genes in the RpoE regulon are more highly expressed in a  $\Delta hfq$  mutant due to activation of the ESR (Figueroa-Bossi *et al.*, 2006), thus it may be speculated that increased levels of MicA and RyeF in the  $\Delta hfq$  mutant result from increased activity of RpoE. There is also a higher proportion of non-Hfq-bound sRNAs among the group of sRNAs which show increased levels in the  $\Delta hfq$  mutant (approximately 70% of up-regulated sRNAs are not associated with Hfq, compared to approximately 30% of down-regulated sRNAs which are not associated with Hfq), indicative of an indirect effect (Figure 3.13 B).

Transcriptomic analysis of the *hfq* mutant coupled with previously published data regarding the sRNAs which are bound by Hfq throughout growth has identified a core group of 46 sRNAs that require Hfq for their activity and stability at ESP. There are many uncharacterised sRNAs in the core group, which are good candidates for further investigation. These core sRNAs are likely to function as canonical *trans*-acting sRNAs, under invasion-associated conditions.



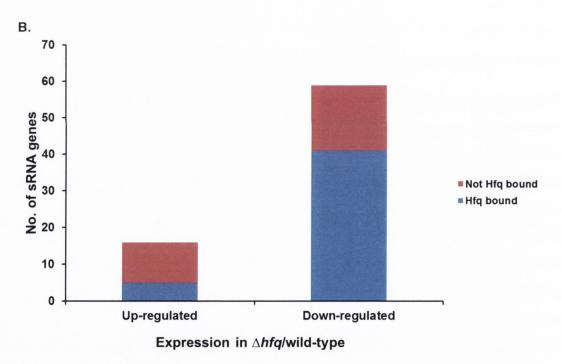


Figure 3.13 Analysis of the sRNA regulon of Hfq

**A.** Venn diagram comparing sRNA genes differentially expressed in  $\Delta hfq$  compared to wild-type, and sRNAs which are bound by Hfq under any condition, or specifically at ESP. **B.** Bar chart demonstrating the number of sRNA genes that are up- or down-regulated in the  $\Delta hfq$  mutant in the context of association with Hfq. Association with Hfq was determined by co-immunoprecipitation (Chao et al., 2012).

#### 3.2.14 Analysis of the sRNA regulon of PNPase

Ribonucleases are considered to be important post-transcriptional regulators of gene expression as ribonuclease functions are important for the processing, maturation and degradation of RNA transcripts. As previously discussed, PNPase is a 3' to 5' exoribonuclease that is a component of the RNA degradosome. PNPase targets both sRNA and mRNA species (Viegas & Arraiano, 2008). The absence of the PNPase protein, therefore, leads to changes in stability of PNPase target genes, rather than changes in transcription. However, RNA-seq, and subsequent gene expression analysis, assess the levels of transcripts in the cell but cannot distinguish between transcripts that are being actively transcribed and those being degraded. Therefore, our data reports on the steady-state levels of the sRNA transcripts, and we cannot comment upon the differences in transcript stability in the absence of PNPase. This will be further discussed in section 3.3.3.

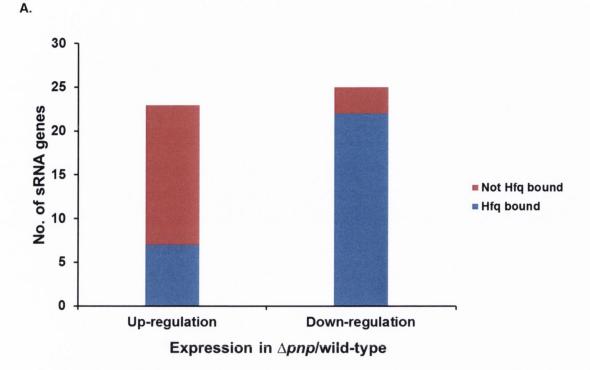
The abundance of 17% (48 genes) of S. Typhimurium sRNA genes is modulated by a factor of 2.5-fold or greater in a mutant strain that lacks PNPase (Figure 3.14 A). Surprisingly, the number of genes which are less abundant is approximately equal to the number of genes which are more abundant following deletion of the pnp gene. A recent study in E. coli found loss-of-function mutations in the pnp gene caused a decrease in stability of RyhB, SgrS and CyaR (De Lay & Gottesman, 2011). The authors suggest that PNPase is a key positive regulator of sRNA stability and function, and the high number of sRNA genes which decrease in abundance in this study agrees with the hypothesis. In this study, the level of the cyaR transcript decreases 2.8-fold in the  $\Delta pnp$  mutant compared to wild-type, while the sgrS transcript decreases 2.3-fold. There is no significant decrease in transcript levels of ryhB-1 or ryhB-2, but the reported effect of PNPase on RyhB may be species-specific and confined to E. coli, or may be a growth phase-specific effect. The De Lay study was carried out during exponential phase while the  $\Delta pnp$  mutant analysis in this study was carried out at ESP. Interestingly, the protective effect of PNPase is restricted to sRNAs, as 96% (592 genes) of protein-coding genes which are differentially expressed in the absence of PNPase are up-regulated, and only 23 protein-coding genes are down-regulated in the  $\Delta pnp$  mutant (Figure 3.14 B).

PNPase co-purifies in a complex with Hfq (Mohanty *et al.*, 2004) and Hfq associates with the RNA degradosome (Vogel & Luisi, 2011), thus we analysed the  $\Delta pnp$  mutant data in

the context of the  $\Delta hfq$  mutant transcriptomic data and Hfq co-immunoprecipitation data (Chao et al., 2012). Among the group of sRNAs that are decreased in abundance in the absence of PNPase, 88% (22 genes) of sRNAs are bound by Hfq in at least one of the conditions used by Chao *et al* and 90% (20 genes) of these Hfq-bound sRNAs are down-regulated in the  $\Delta hfq$  mutant (Figure 3.14 A). The study by De Lay and Gottesman found that the de-stabilising effect of the pnp mutations on sRNA genes were independent of the level and activity of Hfq protein. Instead, it was suggested that PNPase usually protects Hfq-bound sRNAs by limiting the access of RNaseE to the catalytic site of these sRNAs, thereby preventing premature degradation of sRNAs prior to target recognition (De Lay & Gottesman, 2011). The large overlap between sRNAs that are decreased in abundance in the absence of PNPase and in the absence of Hfq, in this study, with sRNAs that co-immunoprecipitate with Hfq, agrees with the hypothesis that PNPase protects Hfq-bound sRNAs from degradation.

The remainder of sRNAs which are affected by the  $\Delta pnp$  mutation accumulate in the absence of PNPase. Only 30% (7 genes) of the up-regulated sRNAs are bound by Hfq under any condition and only 2 genes are bound by Hfq at ESP. Sixteen sRNAs (70%) which are more abundant in the  $\Delta pnp$  mutant are not bound by Hfq and the majority (81%) of these non Hfq-bound sRNAs are also not affected in the  $\Delta hfq$  mutant. The sRNAs which are subject to degradation by PNPase are not Hfq-associated sRNAs, an observation which also agrees well with the study by De Lay and Gottesman. PNPase is the main exoribonuclease involved in degrading transcripts of sRNAs that are normally bound to Hfq, as the sRNAs cycle off or dissociate from the Hfq protein (Andrade *et al.*, 2012). Our data reflect that sRNAs which do not bind Hfq are targets for degradation by PNPase, most likely due to the lack of protection that Hfq normally affords its sRNA targets.

Most of the changes in sRNA gene expression observed in the mutant strain lacking PNPase in this study can be attributed to the role of PNPase in post-transcriptional regulation of the sRNAs, however, it is possible that some effects may be transcriptional. PNPase associates with other proteins (Viegas & Arraiano, 2008) that may affect expression of sRNAs at the level of transcription or activity. There are currently no co-immunoprecipitation data available for the identification of PNPase-associated RNA transcripts, thus it cannot be ruled out that some of the observed changes in gene expression in the absence of PNPase are as a result of indirect transcriptional effects.



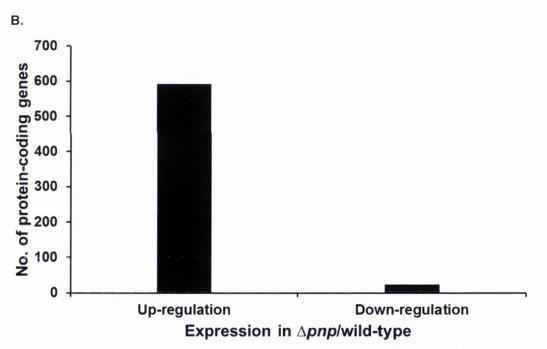


Figure 3.14 Analysis of the PNPase sRNA regulon

**A.** Bar chart showing the number of sRNA genes that accumulate to higher or lower levels in the  $\Delta pnp$  mutant, compared to wild-type. These data are analysed with respect to association with Hfq. Association with Hfq was determined by co-immunoprecipitation (Chao *et al.*, 2012). **B.** Bar chart showing the number of protein-coding genes that accumulate to higher or lower levels in the  $\Delta pnp$  mutant, compared to wild-type, in this study.

#### 3.3 Discussion

To our knowledge, the work presented here is the first large scale transcriptomic study that has been carried out in a panel of *S*. Typhimurium mutants using high throughput next generation sequencing. The use of RNA-seq technology has allowed us to look with unprecedented depth at complex regulons and regulatory mechanisms. In particular, the choice of RNA-seq rather than a microarray-based technology has allowed us to focus, in an unbiased way, on newly identified small non-coding RNAs and to identify the protein factors involved in the regulation of expression of these sRNAs.

#### 3.3.1 Regulation of transcription of S. Typhimurium sRNAs

The general pattern of differential sRNA gene expression in the panel of TF and  $\sigma$ -factor mutants closely mirrors regulatory patterns of protein-coding genes within the cell. There are no apparent differences between sRNA promoters and the promoters of protein-coding genes, indicating that transcription of sRNAs is likely to be regulated by similar mechanisms to those that regulate transcription of protein-coding genes (Gottesman, 2005). It has been estimated that the majority of genes in E. coli are regulated, directly or indirectly, by sRNAs (Holmqvist et al., 2012), and the number of sRNA-mediated regulatory interactions is likely to be similar for S. Typhimurium. Some regulatory proteins such as RpoS and Fur have large regulons of non-coding genes, while other regulators such as the PhoB/R TCS have a more restricted regulon, likely reflecting a more specialised role within the cell. Almost all S. Typhimurium sRNA genes were expressed under the conditions used in this study and over 60% of the sRNA genes are differentially expressed in one or more of the mutant strains, reflecting the importance of this panel of TFs in controlling cellular transcription. Approximately 40% of all differentially expressed sRNAs are bound by the RNA chaperone Hfq, and these sRNAs could play a regulatory role within the cell by binding a target in trans by a classical base-pairing mechanism and contributing to the regulon of their transcription factor.

#### 3.3.2 Regulatory families of sRNA genes

The differentially expressed sRNAs were classified based on their patterns of regulation. Based the hypothesis of "guilt by association", co-regulated or co-expressed genes should share functionality (Quackenbush, 2003). Therefore, correlative analysis to compare the patterns of regulation of sRNA genes to archetypical genes from well-established regulons can provide clues about the functions of the sRNAs, and the regulatory families to which the sRNAs belong. This type of analysis can be applied to any well-established regulon, that contains archetypical genes with a well-defined expression pattern, but the method was used in this work to define a set of SPI1- and SPI2-like sRNAs. This group of sRNAs have regulatory input from either four SPI1-associated or six SPI2-associated regulators (Table 3.2).

Based on the pattern of regulation of the SPI1-like and SPI2-like sRNA genes, we can speculate that the SPI1-like sRNAs may play a role in intestinal epithelial cell invasion or in surviving the host gut environment and that the SPI2-like sRNAs may be involved in survival and replication within the macrophage environment. The absence of most of these sRNAs outside of the *Salmonella* genus, and outside of the *Salmonella enterica* species in the case of the SPI2-like sRNAs, is consistent with these sRNA genes being important for the pathogenicity of *Salmonella*. The genes may have been lost from *S. bongori* and other enteric bacteria as they diverged from their last common ancestor or they may have been acquired by *S. enterica* since the divergence. In either case, I suggest that these sRNAs provide an important and possibly niche-specific fitness benefit to *S.* Typhimurium and other serovars of *S. enterica*. The fact that these sRNAs are not encoded on the pathogenicity islands with which they are associated is an indication that horizontally acquired regulators such as HilD and SsrB have co-opted the function of these sRNAs into their regulons.

The availability of consensus binding motifs for some regulators, such as SsrB, can then allow us to predict which genes may be subject to direct transcriptional regulation by that TF. Two sRNAs, STnc3020 and STnc3090, which show a SPI2-like pattern of expression, contain putative SsrB binding sites within the promoters (Figure 3.9 B). Both sRNAs are highly conserved in *S. enterica*. The SPI1-encoded STnc3020 is conserved in *S. bongori* and the potential for direct binding of SsrB to the STnc3020 promoter region presents the interesting possibility that STnc3020 provides a method of cross-talk

between SPI1 and SPI2. Investigating the differences in expression of STnc3020 between *S. enterica* and *S. bongori* would provide interesting insight into the co-option of sRNAs by TFs to fulfil specific functions. Interestingly, STnc3090 appears to be encoded on an ancestral part of the chromosome (McClelland et al., 2001) and may therefore represent an example of a core gene regulated by a horizontally-acquired regulator.

#### 3.3.3 Regulation of the steady-state levels of S. Typhimurium sRNAs

In addition to investigating the transcriptional regulation of sRNAs in *S.* Typhimurium using RNA-seq, the post-transcriptional regulation of the steady state levels of sRNA transcripts in isogenic mutants of the RNA chaperone Hfq and the exoribonuclease PNPase was also investigated.

sRNAs that bind their targets in trans and contain short imperfect seed sequences for target recognition typically require the RNA chaperone protein Hfq for stability and to aid in target binding. While most of the trans-acting sRNAs characterised in E. coli and S. Typhimurium to date have required Hfq for their function, Hfq-independent sRNA regulation of mRNA targets in trans has been reported in other bacterial species (Wagner, 2013) and similar mechanisms may also exist in Salmonella. It has been demonstrated in E. coli that different sRNAs have differing affinities for Hfq and in some cases, if conditions are optimal and the sRNA-mRNA duplex is sufficiently stable, the requirement for Hfq can be by-passed (Soper et al., 2010). Cis-acting sRNAs and sRNAs which have protein, rather than mRNA, targets typically do not require Hfq for their function (Vogel & Luisi, 2011). In addition, as has been previously mentioned (section 1.3.3 and 3.2.12), the interaction between sRNAs, their mRNA targets and Hfq can be dynamic and transient, resulting in periods of time that the sRNAs are not associated with Hfq (Wagner, 2013), while competition for limiting amounts of Hfq protein and over-expression of Hfq-bound sRNAs can also affect the binding properties of some sRNAs to Hfq (Moon & Gottesman, 2011, Papenfort et al., 2009). The competitive binding to Hfq between sRNAs makes it unlikely that sRNAs will be Hfq-bound under conditions where regulation of their target genes is not required. Therefore some of the sRNAs which have been classified as "not bound" by Hfq may be associated with Hfq under certain conditions, or may regulate target genes via an unknown Hfq-independent mechanism. Nevertheless, Hfq remains the most important factor for regulating function

and stability of sRNAs (Vogel & Luisi, 2011) and approximately 80% of sRNAs which are differentially expressed in the absence of Hfq are Hfq-associated under at least one environmental condition (Figure 3.13 A).

Surprisingly, an equal proportion of sRNA transcripts are specifically down-regulated and up-regulated in the absence of PNPase (Figure 3.14 A), a feature which is not true of protein-coding genes, which generally accumulate to a higher level in the absence of PNPase (Figure 3.14 B). The majority of sRNA genes which accumulate to a higher level as a result of the pnp mutation are not bound by Hfq, refecting the fact that PNPase is a key factor in degrading non-Hfq associated sRNAs (Andrade et al., 2012). It would be interesting to investigate the stability of up-regulated sRNAs in a double  $\Delta hfq\Delta pnp$ mutant, as Andrade and co-workers observed that a pnp mutation restored expression of micA, rybB, sgrS and glmY to wild-type levels in the  $\Delta hfq$  mutant. Hfq protects sRNA transcripts, under normal conditions, but sRNA transcripts are more vulnerable to degradation by PNPase in the absence of Hfq, thus the absence of PNPase results in reduced sRNA degradation. The sRNA genes which are down-regulated in the absence of PNPase are overwhelmingly Hfq-bound sRNAs (Figure 3.14 A). A PNPase-mediated protective effect has been documented for sRNAs in E. coli previously and the de-stabilisation of ryhB, sgrS and cyaR in a  $\Delta pnp$  mutant was found to be independent of levels or activity of the Hfq protein (De Lay & Gottesman, 2011). It was suggested that the protective effect of PNPase was indirect, caused by PNPase modulating the access of RNaseE to Hfq-bound sRNAs prior to target recognition.

Differences in growth condition may account for the apparently antagonistic roles played by PNPase in the cell. PNPase-mediated RNA degradation is growth-phase dependent and is more important in stationary phase rather than exponentially growing cultures (Andrade et al., 2012). The majority of studies that have shown PNPase to be a key factor in degrading sRNAs were conducted in late stationary phase cultures (Andrade et al., 2012, Andrade & Arraiano, 2008, Viegas *et al.*, 2007), while the protective effect of PNPase was demonstrated during exponential growth. The work presented here was carried out at early stationary phase and the intermediate growth phase may account for differences in the effects of the *pnp* mutation.

It is also important to note that, as an exoribonuclease, PNPase is generally unable to digest extensive secondary structure and usually requires an endonuclease to initiate the

digestion of RNA transcripts (Cheng & Deutscher, 2005). For this reason, the absence of the PNPase exoribonuclease can result in the accumulation of decay intermediates (Viegas et al., 2007). RNA-seq involves the sequencing of cDNA fragments and the sequencing reads are then mapped to a reference genome (Ozsolak & Milos, 2011). Algorithms, such as TPM (section 2.6.9), are used to calculate gene expression values based on the reads mapped to a particular gene (Wagner et al., 2012). TPM values may not reflect a change in full length transcripts, if a similar number of reads map to decay intermediates within the gene. Care must, therefore, be taken when analysing gene expression values from mutant strains that lack ribonucleases. Changes in sRNA processing between wild-type and the  $\Delta pnp$  mutant can be detected using sequencing data visualisation programmes such as IGB or jbrowse, however.

The use of RNA-seq to detect differential gene expression in the strains that lack Hfq and PNPase provides different information than the use of RNA-seq for detection of differential gene expression in each of the TF mutants. The Hfq and PNPase proteins do not act at the level of sRNA transcription. Instead, Hfq binds sRNAs to enhance their activity and to protect sRNA transcripts from degradation by cellular ribonucleases (Vogel & Luisi, 2011). PNPase, as a key component of the RNA degradosome, is primarily involved in 3' to 5' exoribonuclease activity (Viegas & Arraiano, 2008). Thus, RNA-seq technology can be used to detect changes in transcript levels, resulting from increased or decreased stability of the transcript, rather than the detection of changes in transcript initiation, resulting from perturbations of transcriptional networks. Use of the RNAP-inhibitor antibiotic, Rifampicin, to halt cellular transcription and analysis of cellular transcripts over time, following the inhibition of transcription, would complement the Hfq and PNPase transcriptomic data. The inhibition of de novo transcription allows us to distinguish between RNA steady state levels and RNA turnover, giving a more comprehensive view of the direct sRNA targets of Hfq and PNPase. Abundantly transcribed sRNAs may appear not to be affected in the mutant strains if cellular transcription continues at the same rate in the wild-type and mutant strains, however the half-life of these abundant transcripts could be affected. The transcriptomic experiments, in the absence of Rifampicin, provide a useful method to detect which sRNA genes are most highly affected by each mutation, and to report on the steady state levels of sRNAs in each mutant strain compared to wild-type.

## 3.3.4 The benefits and limitations of an RNA-seq-based approach for the investigation of bacterial regulons

RNA-seq provides an ideal method for studying gene expression at single nucleotide resolution. Large amounts of data can be generated rapidly and easily, relative to more laborious protocols such as DNA microarrays (Ozsolak & Milos, 2011). Extraction of good quality total or depleted RNA is important but most sequencing facilities offer cDNA library preparation services to ensure consistent high quality samples for sequencing. Sequencing technologies themselves are improving and advancing regularly, resulting in the generation of large amounts of good quality data, within a short timescale. RNA-seq technology is also becoming more accessible to more researchers, meaning that high throughput next generation sequencing and RNA-seq are contributing to the expansion of our knowledge in almost all fields of biology. The development of tools for bioinformatics and data handling programmes permit rapid and convenient downstream analyses, allowing us to distil the important information from the, almost overwhelming, amount of data that can be produced using RNA-seq (Mutz et al., 2013). We have shown that the use of RNA-seq to determine TF regulons provides accurate information that agrees closely with what has been previously shown using transcriptional fusions or microarrays, but the increased sensitivity and dynamic range of RNA-seq allows us identify more subtle changes in gene expression, as well as to detect expression of previously unannotated non-coding RNA species, thereby increasing our understanding of complex regulatory pathways.

However there are some limitations to the approach used in this study. Firstly, RNA is extracted from a population of bacteria for use in RNA-seq, providing an average view of the transcriptomes of millions of cells. However, it is becoming increasingly clear that genetically identical cells can often be phenotypically different and that sub-populations of cells or outlier cells play important roles in disease outcome or antibiotic tolerance (Helaine *et al.*, 2014, Diard *et al.*, 2013). Developments in the sensitivity and automation of the instruments necessary for isolating and sequencing RNA from single cells has led to great advances in the field of eukaryotic gene expression, and single-cell RNA-seq of prokaryotic cells is also becoming a possibility (Saliba *et al.*, 2014). Single-cell transcriptomic analysis will allow the detection of even more subtle, and often biologically significant, changes in gene expression that could be masked at the level of whole bacterial populations.

Furthermore, RNA-seq of mutant strains and comparison of changes in gene expression does not distinguish between direct and indirect regulation. The complexity of bacterial regulatory networks and the interplay between different regulators means that many of the putative interactions that are detected in the absence of a TF are as a result of perturbations in regulatory cascades rather than direct control of gene expression. Binding experiments, such as electrophoretic mobility shift assays (EMSA), were traditionally used to demonstrate that a protein could directly bind a specific fragment of DNA and more recently Chromatin Immunoprecipitation (ChIP) has become the method of choice to prove direct protein binding *in vivo* (Cai & Huang, 2012). ChIP followed by qPCR can be used to determine protein binding within a particular sequence of interest, while ChIP-on-chip technology involves the hybridisation of protein-associated DNA to a microarray to determine all DNA binding sites of a particular protein.

However ChIP-on-chip has more recently been superceded by ChIP-seq which involves the use of deep sequencing technology to determine all DNA molecules bound by DNA-associated proteins (Cai & Huang, 2012). As previously discussed, sequencing-based approaches, rather than microarray-based technology, has increased sensitivity and reduced bias. The use of ChIP-seq also allows for the identification of transcription factor binding site sequences at the level of the individual nucleotide. Ideally studies can combine gene expression data obtained from RNA-seq experiments comparing wild-type and regulatory mutants with ChIP-seq data to relate the changes in gene expression to binding and direct regulation by those regulators. Studies of that nature, however, are prohibitively expensive on a large scale, for example, with a panel of regulatory mutants, such as those presented in this work. Instead, the use of the RNA-seq-based gene expression data from the panel of mutants allows us to identify regulatory patterns of genes across a number of conditions and strains which provides a good prediction of the regulatory pathways involved in controlling expression of these genes. Motif analysis can supplement this information to predict direct protein binding sites and provide focussed data on highly likely candidate gene targets. ChIP-qPCR can then be used as a cheaper alternative to large-scale ChIP-seq to validate these candidate interactions.

The use of RNA-seq to investigate the transcriptomes of important bacterial pathogens, such as *S*. Typhimurium, gives us important clues about gene expression and regulatory interactions but important findings must be validated using the molecular biological

approaches previously described. It has become evident that we need complete descriptions of biological molecular networks, including all components, interactions and environmental inputs, providing a better understanding of the inherent complexity of biological systems than the investigation of individual components of biological circuits can provide (Hebrard et al., 2011, Golubeva et al., 2012). This complete description is a long term aim that will be achieved through the integration of information from a number of datasets, including transcriptomics, epigenomics, proteomics, global ChIP studies and The integration of such diverse data will require careful analysis and good bioinformatics tools to make the information easily accessible to the scientific community (Hebrard et al., 2011). Great strides have been taken towards achieving the aim of integrated systems biology analysis of the model organism E. coli gene regulation. The online resources RegulonDB, EcoCyc and GenExpDB provide curated information on E. coli gene expression from thousands of transcriptomic experiments (Hebrard et al., 2011). However, despite over 7 decades of research, the complexities of the E. coli genome are only becoming realised in recent times with the advent of RNA-seq technologies (Conway et al., 2014). Thus, a lot remains to be done to achieve the aim of using integrated approaches to attain a clear picture of the multi-factorial process involved in the pathogenesis of infection. Our hope is that this transcriptomic analysis of S. Typhimurium virulence regulons will contribute to the "big picture" of S. Typhimurium infection.

### **Chapter 4**

# Investigating the transcriptional regulation of the SPI1-like sRNA, STnc520

#### 4.1 Introduction

#### 4.1.1 SPI1-associated sRNAs

STnc520 was characterised as a SPI1-like sRNA based on the pattern of STnc520 expression across a panel of 18 regulatory mutants under 5 environmental conditions (Figure 3.2.10). The STnc520 regulatory pattern is similar to that of InvR and DapZ, both of which are dependent on HilD for transcription (Chao et al., 2012, Pfeiffer et al., 2007). InvR was first identified biocomputationally in a screen for orphan pairs of  $\sigma^{70}$ promoters and p-independent terminators in IGRs. InvR is particularly interesting as it is encoded within the SPI1 gene cluster, downstream of the *invH* gene (Pfeiffer et al., 2007). InvR is induced at ESP, in a low oxygen environment and at high osmolarity, conditions which induce expression of SPI1 genes (Ellermeier & Slauch, 2007). InvR is a highly stable sRNA (half-life >30 minutes), which is completely dependent on Hfq for its stability. InvR transcription is activated by HilD, and deletion of other SPI1-encoded or SPI1-associated regulators does not affect InvR expression. Furthermore, EMSAs were used to demonstrate that HilD directly binds the invR promoter region (Pfeiffer et al., 2007). However, InvR does not regulate expression of SPI1 genes or play a role in SPI1 secretion. Rather, STnc520 represses translation of the major outer membrane porin ompD mRNA, providing a good example of a horizontally-acquired element regulating expression of a gene encoded as part of the core ancestral Salmonella chromosome. It has been suggested that the InvR-mediated regulation of ompD may have evolved to maintain envelope stability to allow establishment of the membrane-anchored SPI1 translocation apparatus (Pfeiffer et al., 2007).

DapZ, on the other hand, is not encoded on a pathogenicity island but is located at the 3' end of an ancestral gene, dapB. DapZ co-immunoprecipitates with Hfq and requires Hfq for stability. DapZ accumulation peaks at ESP, suggesting co-regulation of DapZ with SPI1 genes (Chao et al., 2012). Investigation of dapZ expression, in mutants lacking the major SPI1-encoded and SPI1-associated TFs, show that HilD is the most important TF for dapZ transcription. Ectopic expression of HilD induces expression of DapZ in heterologous systems that lack other SPI1-associated TFs. As with InvR, DapZ does not control expression of SPI1 or invasion-associated genes and instead, targets operons which encode major ABC transporter systems oppA and dppA, providing a mechanism to

repress transport of oligopeptides and dipeptides, respectively, under invasion conditions (Chao et al., 2012).

A third sRNA, IsrJ, is also regulated by a SPI1-encoded TF. Expression of the island-encoded sRNA, IsrJ, was found to be reduced, but not completely abrogated, in the absence of HilA. However, unlike InvR and DapZ, IsrJ appears to play a more direct role in virulence as a  $\Delta isrJ$  mutant strain demonstrates reduced invasion rates in HeLa cells. Furthermore, decreased translocation of SPI1 effector proteins was detected in the absence of IsrJ (Padalon-Brauch et al., 2008). IsrJ was not classified as a SPI1-like sRNA in this study, however, as isrJ expression does not decrease in a  $\Delta hilD$  mutant (Appendix III). There appears to be one or more additional factors involved in the regulation of isrJ transcription as IsrJ is still expressed, albeit to a lower level, in a  $\Delta hilA$  mutant (Padalon-Brauch et al., 2008), and this possible regulatory redundancy may explain the lack of an effect of deleting hilD on IsrJ expression, in this study.

The island-encoded sRNA IsrM is important for Salmonella invasion and survival within macrophages (Gong et al., 2011). Expression of isrM is highly induced under conditions that mimic those present during the early stages of intestinal infection, namely low pH, high osmolarity and low oxygen. IsrM is also important in vivo as IsrM expression is highly induced in Salmonella harvested from the ileum and spleen of infected BALB/c mice (Gong et al., 2011). Mutation of isrM results in dysregulation of a number of SPI1 effector proteins. This global dysregulation of Salmonella effector proteins is thought to be mediated by IsrM targeting and repression of hilE mRNA, which encodes a protein that represses SPI1 via HilD (Baxter et al., 2003, Gong et al., 2011). IsrM also targets and represses translation of the SPI1 effector-encoding sopA mRNA (Gong et al., 2011). Mutants lacking IsrM are attenuated in BALB/c mice, and survive for up to 15 days post infection, while mice infected with wild-type or a  $\Delta isrM$  mutant complemented with plasmid-encoded IsrM are killed within 5 days post-infection (Gong et al., 2011). It is unclear if isrM expression is controlled by a SPI1-associated regulator. IsrM is not expressed in the wild-type or mutant strains used in this study (Appendix I), under 22 environmental conditions (Kröger et al., 2013) or within murine macrophages (Srikumar et al., 2014) in S. Typhimurium 4/74. Gong et al investigated IsrM in S. Typhimurium 14028, and the differences in the IsrM expression profile may reflect strain differences between 4/74 and 14028.

The association of these sRNAs (InvR, DapZ, IsrJ and IsrM) with SPI1, either through SPI1-encoded regulatory proteins or via sRNA-mediated regulation of SPI1 function, demonstrates that the SPI1 locus is an important hub for sRNA-mediated gene regulation. This observation is confirmed by the phenotypic link between the RNA chaperone, Hfq, and SPI1 gene function. As previously discussed (section 1.3.2), a  $\Delta hfq$  mutant is attenuated for adherence to and invasion of cultured HeLa cells, and shows an altered SPI1 protein secretion profile (Sittka et al., 2007). These effects were later found to be due direct association of Hfq with genes encoded within SPI1, including the primary SPI1 activator, HilD (Sittka et al., 2008). Therefore, the identification of another sRNA, STnc520, which demonstrates a SPI1-like pattern of expression and is dependent on HilD for transcription merited further investigation to understand the role of STnc520 in the context of the SPI1 regulon.

#### 4.1.2 SprB

Five TFs are encoded on the SPI1 pathogenicity island. HilD acts at the top of the SPI1 regulatory cascade and is part of a feed-forward loop with the SPI1-encoded HilC, and RtsA (which is encoded outside of SPI1 (Ellermeier et al., 2005)). HilA is the central SPI1 regulator, which mediates expression of the majority of genes encoding proteins involved in the structural apparatus of the TTSS and SPI1 effector proteins, both directly and indirectly through HilA activation of the gene encoding the regulatory protein, InvF (Thijs et al., 2007) (Figure 1.3).

The final SPI1-encoded regulator, SprB, is less well characterised than the other SPI1-encoded TFs. SprB was first identified and characterised as a regulatory protein, based on its homology to the LuxR family of Helix-Turn-Helix proteins, shortly after the discovery of SPI1 (Eichelberg *et al.*, 1999). However, a role for SprB in regulating SPI1 gene expression was not identified, as loss-of-function mutations or constitutive over-expression of SprB do not affect *Salmonella* invasion of cultured intestinal epithelial cells and do not affect expression of SPI1-associated effector proteins, encoded on SPI1 or elsewhere on the chromosome. However SprB expression was induced by the same environmental stimuli that trigger expression of other SPI1 genes suggesting that SprB is functionally associated with SPI1 but may target unknown TTSS substrates (Eichelberg et al., 1999).

SprB was later identified as the molecular link between SPI1 and SPI4, which encodes a Type 1 secretion system that produces a non-fimbrial adhesion. Activity of the *siiA* promoter is reduced in the absence of SprB and SprB rescues *siiA* promoter activity in a ΔSPI1 mutant. A co-precipitation assay was used to demonstrate that SprB directly binds the *siiA* promoter (Saini & Rao, 2010). HilA binds and directly activates expression from the *sprB* promoter, while SprB has a slight repressive effect on SPI1 gene expression, via *hilD*, following static growth for 12 hours (Saini & Rao, 2010). These data indicate that SprB does play an important role during invasion, by mediating expression of a non-fimbrial adhesin which has been shown to make contact with the apical side of polarised intestinal epithelial cells, a process which is necessary for SPI1-mediated translocation of effector proteins and invasion (Gerlach *et al.*, 2008). The regulatory link between SPI1 and SPI4 is, therefore, an important one. In addition, SprB may be involved in repressing expression of SPI1 genes later in infection (Saini & Rao, 2010).

#### 4.1.3 SPI11

There are 13 SPIs in the genome of S. Typhimurium 4/74 (Kröger et al., 2012). The SPI11 locus encodes a number of virulence-associated genes, including msgA and the PhoP-dependent pagC and pagD, all of which are important for bacterial virulence and macrophage survival (Gunn et al., 1995). The pliC gene which encodes a secreted periplasmic lysozyme inhibitor protein is encoded on SPI11, as well as genes encoding membrane spanning lipoproteins, EnvE and EnvF, the cold shock protein CspH and a predicted bacteriophage protein STM1239. In addition to the protein-coding genes, there are three sRNA genes (IsrC, STnc3420 and STnc520) and one tRNA gene (STM1247 (tRNA-Arg)) present within the 4/74 SPI11 locus (Kröger et al., 2013). IsrC overlaps the 3' end of the msgA gene and both RNAs undergo mutual degradation when expressed in cis from a multi-copy plasmid. IsrC is highly expressed under a number of shock conditions such as oxidative shock, and expression of IsrC is highly-induced early during infection of macrophages but declines as the infection progresses (Padalon-Brauch et al., 2008). In addition to the 3 sRNAs annotated in 4/74, a 4<sup>th</sup> sRNA gene has been identified on SPI11 in S. Typhimuirum serovar UK1. This sRNA is located in the IGR between cspH and envE, and is named RaoN (RNA involved in response to acid and oxidative stress during nutrient limitation) (Lee et al., 2013). RaoN is most highly induced in

response to a combination of nutrient limitation and oxidative stress. The  $\Delta raoN$  mutant demonstrates reduced levels of intracellular replication within murine macrophages, compared to the wild-type strain and a complemented  $\Delta raoN$  mutant strain. RaoN targets the ldhA gene, which encodes a lactate dehydrogenase, and fine-tuning of the expression of ldhA appears to be important for intra-macrophage survival (Lee et al., 2013). Both IsrC and RaoN appear to be important during conditions which mimic the intra-macrophage environment, similar to PagC, PagD and MsgA. The presence of a gene encoding a lysozyme inhibitor and genes involved in intra-macrophage survival makes it tempting to speculate that the overall role of SPI11 in pathogenicity is to resist mechanisms of the host innate immune system to allow Salmonella to withstand the stressful intracellular environment.

The protein-coding genes of SPI11 are highly conserved in all *S. enterica* serovars, with the exception of serovar Paratyphi B Java which does not contain any SPI11 genes. The *envF* gene is also missing in some *S. enterica* serovars, such as Typhi and Paratyphi A, as well as more distantly related *S. enterica* strains, such as those belonging to subspecies *indica*, *salamae*, *houtenae*, *diarizonae* and arizonae. The pagD gene is also not present in strains from other subspecies (Desai et al., 2013). Most of the island, with the exception of *envF* and pagD, is conserved in *S. bongori* while there is very little conservation of protein-coding SPI11 genes outside of the *Salmonella* genus. Despite this, msgA is conserved in *Citrobacter* and *Yersinia* species, cspH is conserved in non-pathogenic and pathogenic strains of *E. coli* and pagC is conserved in *Enterobacter cloacae* (Desai et al., 2013). The presence of a complete and intact SPI11 in *Salmonella enterica* subspecies I only (i.e. subspecies *enterica*) suggests that this pathogenicity island plays a niche-specific role in human infection and infection of other warm blooded hosts.

#### 4.1.4 Formate metabolism

Formate is a short chain fatty acid (SCFA) which is produced during anaerobic metabolism by enteric bacteria. There are high levels of formate in the human intestine as a result of the utilisation of polysaccharides by the resident microbiota of the gut. Other SCFAs, such as acetate, propionate and butyrate are also produced in this way and there are distinct patterns of location of each SCFA within the gut. High levels of acetate and formate in the distal small intestine (the ileum) have been suggested to act as signals

to induce *Salmonella* invasion in this area (Lawhon et al., 2002, Huang *et al.*, 2008), while butyrate and propionate are produced at higher levels in the large intestine (the caecum and colon) and the latter SCFAs serve to repress *Salmonella* invasion gene expression at these sites (Gantois *et al.*, 2006, Lawhon et al., 2002). In this way *Salmonella* makes use of environmental signals to ensure temporally and spatially appropriate gene expression. Metabolic by-products of formate oxidation are not important for formate-mediated induction of SPI1 genes, as mutations in the genes encoding the key formate dehydrogenases do not affect invasion gene expression (Huang et al., 2008).

Formate is a key metabolite in energy metabolism in many bacteria, including Salmonella and other enterics, and formate can be oxidised during aerobic and anaerobic metabolism. anaerobic metabolism formate is produced from pyruvate by the pyruvate-formate-lyase (Pfl) protein. At a neutral pH, formate is excreted from the bacterial cell. However, if nitrate is present under acidic conditions, nitrate-inducible formate dehydrogenase-N (FDH-N, encoded by fdnGHI) oxidises formate, and formate acts as an electron donor for nitrite and other substrates of anaerobic respiration. If nitrate is not present under acidic conditions, formate dehydrogenase-H (FDH-H, encoded by fdhF) becomes induced and prevents further acidification of the cell by converting formate to CO<sub>2</sub> and H<sub>2</sub>. Formate is present at low levels under aerobic conditions as Pfl is inactive and pyruvate is not cleaved to generate formate. In addition, formate dehydrogenase-O (FDH-O, encoded by fdoGHI) is induced under aerobic conditions and oxidises formate via the aerobic respiratory chain (Leonhartsberger et al., 2002). Interestingly, translation of the mRNA of the gene which encodes the central transcriptional regulator of the formate regulon, fhlA, is under the control of the oxidative stress-induced sRNA, OxyS (Altuvia et al., 1997). This is thought to prevent unnecessary production of oxygen sensitive formate metabolic systems during an oxidative burst, such as that experienced in the transition between anaerobic growth in the intestine to the more aerobic macrophage environment.

#### 4.1.5 Sulphur metabolism

Sulphur is an essential element for growth in all organisms and is primarily required as a component of the sulphur-containing amino acids, cysteine and methionine, but has other

secondary roles within the cell. In bacteria, such as Salmonella, sulphur is generally derived from inorganic sulphate via an energy consuming assimilatory pathway, although other sulphur-containing compounds may also act as a sulphur source. Inorganic sulphate is transported into the bacterial cell and is activated by conversion to adenosine 5'-phosphosulphate, followed by 3'-phosphoadenosine-5'-phosphosulphate (PAPS), which is subsequently reduced to sulphite and then sulphide. The reduced sulphur compounds produced by this pathway are used to yield cysteine. The production of cysteine from sulphide is mediated by the presence of O-Acetylserine, which is formed from the amino acid serine and acetyl-CoA (Kredich, 1992). Other sources of sulphur require alternative uptake systems and are converted to cysteine via alternative pathways. The genes of the cysteine regulon encode the proteins that are required for the uptake of sulphur sources such as sulphate, L-cystine, thiosulphate and taurine and for the synthesis of cysteine from these sources. The cysteine regulon requires transcriptional activation by CysB, the inducer molecule N-acetyl-L-serine and sulphur limitation. Good sulphur sources, such as cysteine, repress the sulphur regulon by competing with the inducer molecule for binding to the CysB activation domain, or by repressing the synthesis of inducer molecules. Genes within the sulphur regulon become de-repressed in the presence of poor sulphur sources (Quan et al., 2002).

Maintaining cellular redox homeostasis is essential for bacterial metabolism and cellular integrity. Bacteria have, therefore, developed a number of sensors to monitor the redox state of the cell. The thiol side-chain of cysteine is nucleophilic and easily oxidised into a number of different redox states. This property enables cysteine to act as a redox sensor, making cysteine highly reactive and allowing it to play a role in many biological functions, such as, the formation of disulphide bonds, to provide or to alter protein structure, and metal ion binding, including the formation of iron-sulphur clusters (Green & Paget, 2004). Oxidative damage to wild-type *S.* Typhimurium increases the need for cysteine and involves up-regulation of members of the CysB regulon, while cysteine auxotrophs are constitutively under oxidative stress (Turnbull & Surette, 2010). CysB also has a global effect on other cellular processes, such as carbon metabolism, and affects expression of the formate dehydrogenase-O-encoding *fdo* operon, at the transcriptional level. This cross-talk between regulons likely allows the bacteria to co-ordinate energy consumption and metabolic processes with the available nutrients (Quan et al., 2002).

#### 4.1.6 Copper homeostasis

Copper is an essential trace element for bacteria, but copper can also be toxic and has long been used as an anti-microbial agent (Hodgkinson & Petris, 2012). Copper has the ability to cycle between oxidation states to form either the cuprous ion (Cu(I)) or the cupric ion (Cu(II)). This property allows copper to interact with hydrogen peroxide and to catalyse reactions which form hydroxyl radicals which can damage bacterial cell structures, proteins and nucleic acids. Copper also damages proteins by binding to amino acids, such as cysteine, and excluding the native metal cofactors, such as iron-sulphur clusters, from their binding sites (Hodgkinson & Petris, 2012). Copper accumulates in the macrophage phagosome during Salmonella infection as part of the host immune response (Osman et al., 2010). Salmonella has, therefore, developed methods of resisting copper-mediated bacterial killing and of maintaining copper homeostasis. These methods include copper export, sequestration of copper and oxidation of copper to the less toxic cupric ion. Two related P<sub>1B</sub>-type ATPases, CopA and GolT are necessary for export of copper from the cytoplasm to the periplasmic space during Salmonella infection within the phagosomal environment (Osman et al., 2010). golT (STM0353) expression is under the transcriptional control of GolS (STM0354). copA expression is regulated by the copper-sensing transcriptional regulator CueR, as is expression of the gene encoding the periplasmic copper binding protein CueP (STM3650). CueP mediates Salmonella resistance to copper toxicity in the periplasm, particularly under anaerobic conditions (Pontel & Soncini, 2009). CueP also provides copper for activation of periplasmic superoxide dismutase (SOD) proteins, which are necessary for resistance to ROS, such as those encountered by S. Typhimurium in the phagosome (Osman et al., 2013). CueP acts as an intermediate between the P<sub>1B</sub>-type ATPases and the copper-zinc dismutase protein (SodCII) as CueP binds the exported periplasmic copper and transfers it to SodCII. Accordingly, members of the CueR and GolS regulons provide resistance to macrophage killing and play an important role in virulence.

#### 4.2 Results

#### 4.2.1 Characterisation of STnc520

STnc520 was first identified in S. Typhimurium by co-immunoprecipitation with the RNA chaperone Hfq at ESP (Sittka et al., 2008). STnc520 is 85 nt in length and is encoded on SPI11, in the IGR between the convergently transcribed pliC and pagC genes. There is also a tRNA-Arg (STM1247) gene transcribed on the opposite strand within the same IGR (Figure 4.1 A). However STnc520 was not classified as Hfq-bound in the recent study by Kröger et al which screened for trans-acting regulatory sRNAs of S. Typhimurium, based on a Hfq enrichment factor >5 after co-immunoprecipitation with Hfq at 7 different growth phases, compared to a control co-immunoprecipitation (Chao et al., 2012, Kröger et al., 2013). To further characterise the association of STnc520 with Hfq, the stability of the STnc520 transcript in wild-type and the  $\Delta hfq$  mutant was investigated at ESP, as previously described (section 2.6.10). RNA was extracted at different time-points before and after the addition of Rifampicin and total RNA was probed to detect the levels of the STnc520 transcript by northern blot (Figure 4.1 B). The intensity of bands from 2 independent biological experiments was quantified by densitometry and the transcript half-life was calculated, as described (section 2.6.10) (Figure 4.1 C). The half-life of the STnc520 transcript in wild-type cells is approximately 12 minutes. In the context of the sRNA half-lives that have been reported (Vogel et al., 2003), STnc520 is a relatively stable sRNA. Surprisingly, however, while the steady state levels of STnc520 decrease in the  $\Delta hfq$  mutant, the stability of the transcript does not decrease, in fact the half-life of the transcript appears to be slightly longer in the mutant than in the wild-type strain (approximately 13 or 14 minutes) (Figure 4.1 B and C). Hfq binds and activates expression of hilD mRNA (Sittka et al., 2008), and STnc520 expression decreases in the absence of HilD at ESP (Figure 3.6 A). It is, therefore, possible that the decrease in STnc520 transcript levels in the absence of Hfq (Figure 4.1 B and C) is an indirect effect, due to decreased expression of HilD and a concomitant reduction in STnc520 transcription, rather than increased levels of STnc520 degradation in the absence of Hfq. However, the fact that STnc520 co-immunoprecipitated with Hfq indicates, that Hfq does bind STnc520, although the interaction may be weak and transient. Hfq may be important for STnc520 activity but not for stability of the STnc520 transcript, as has been previously shown for GcvB in E. coli (Moon & Gottesman, 2011).

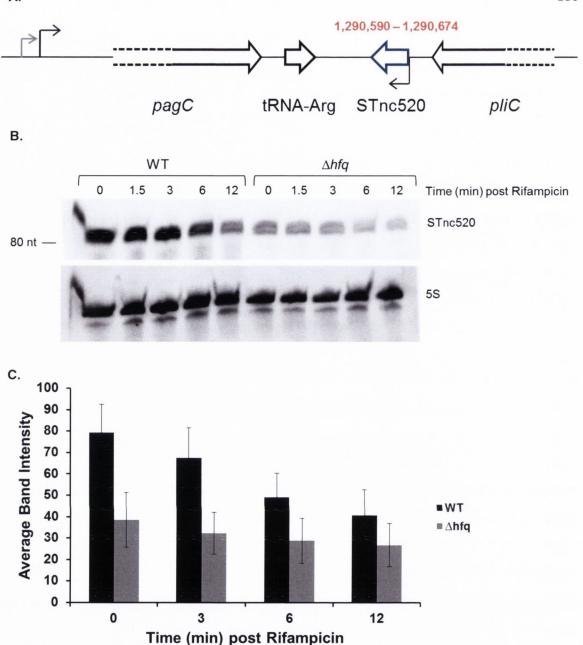


Figure 4.1 Chromosomal location and stability of STnc520

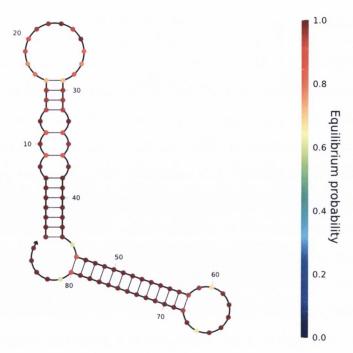
A. Schematic showing the chromosomal location of the STnc520 gene. The STnc520 gene (white arrow with blue outline) is located on SPI11, and is encoded on the minus strand, downstream of the gene encoding the lysozyme inhibitor PliC (white arrow with black outline). Bent arrows denote TSS, black arrows are primary TSS and smaller grey arrows are secondary TSS (Kröger et al., 2013). Arrows indicate the direction of transcription of each gene. Dotted lines indicate that the genes are not shown to scale. The chromosomal co-ordinates of STnc520 are written in red. В. northern blot showing the stability of the STnc520 transcript in wild-type and  $\Delta hfq$  cells following the inhibition of transcription by the addition of Rifampicin, as previously described (section 2.6.10). Numbers correspond to the times (in minutes), before and after addition of Rifampicin, that cells were withdrawn for RNA extraction. ribosomal RNA was probed as a loading control. C. Densitometry measurement based on the intensity of each northern blot band from 2 independent experiments, as measured using ImageJ software. Error bars are based on the standard error from the maximum and minimum intensity reading across a given band in both experiments.

#### 4.2.2 Predicted structure and conservation of STnc520

The stability of STnc520 is likely to reflect the formation of stable secondary structures within the transcript, which protect the sRNA from degradation by ribonucleases. Software available from www.nupack.org (Zadeh et al., 2011) was used to predict the minimum free energy structure of STnc520 (Figure 4.2 A). The predicted structure is coloured according to the probability of these base-pairs, or no base-pairing, occurring, as shown in the colour bar. Based on the probability of paired and non-paired nucleotides of STnc520, the predicted structure is a highly likely structure for STnc520 to adopt. The predicted structure comprises 2 stem-loops that arise from palindromic sequences within the STnc520 gene. The first stem-loop (nt 1-44) is a less stable structure due to 3 internal bulges that contain a number of A-U nucleotides. The second stem-loop (nt 46-80) is a more stable G-C rich region of the sRNA and is a predicted ρ-independent terminator. The stable base-pairing of the stems is likely to protect STnc520 from degradation by the single-strand-specific endonuclease, RNase E, and from 3' to 5' exonucleases, such as PNPase, RNaseR and RNase II, which typically cannot digest stable secondary structures (Viegas & Arraiano, 2008). Therefore, the structure shown in Figure 4.2 A may explain the long half-life of the STnc520 transcript, and the lack of requirement for Hfq for protection from degradation.

Figure 4.2 B is a multiple alignment of the STnc520 sequence from ten enteric bacteria. The 4/74 STnc520 sequence is highly conserved in other *S. enterica* serovars, is approximately 80% conserved in *S. bongori* and is approximately 40% conserved in pathogenic *E. coli*, *Shigella flexneri* and *Citrobacter koseri* (Figure 3.12 A). Alignment of the gene sequence from these bacteria allows the identification of the most highly conserved regions of the gene, which are likely to be functionally-relevant. Nucleotides 60-70 are highly conserved, and constitute the loop of the predicted ρ-independent terminator. The other stem-loop region (nt 1-44) is relatively well conserved, especially within the *Salmonella* genus and particularly in *S. enterica*. As mentioned, the stability of the terminal and internal loop regions is much lower than the stem regions due to the lack of predicted base-pairing. It is, therefore, often these loop regions which are involved in base-pairing interactions with target mRNAs, as the energy required to make loop structures accessible is much lower than the energy required to break stem structures (Tjaden *et al.*, 2006), thus this loop may be a structure of functional importance.





Free energy of secondary structure: -37.40 kcal/mol

B.

	10	20	30	40	50	60	70	80	
S. Typhimurium 4/74	GGCTGGTAAC	CATTC-ATAT	TCGCACTATT	GAACGATCAC	CGGCCAACCG	C-GCCAGGGT	CTTGCATACG	CTCTGGC-GC	GGTTAAAT
S. Typhimurium D23580	GGCTGGTAAC	CATTC-ATAT	TCGCACTATT	GAACGATCAC	CGGCCAACCG	C-GCCAGGGT	CTTGCATACG	CTCTGGC-GC	GGTAAAAT
S. Typhimurium D23580 S. Dublin	GGCTGGTAAC	CATTC-ATAT	TCGCACTATT	GAACGATCAC	CAGCCAACCG	C-GCCAGGGT	CTTGCATACG	CTCTGGC-GC	GGTTAAAT
S. Enteritidis	GGCTGGTAAC	CATTC-ATAT	TCGCACTATT	GAACGATCAC	CAGCCAACCG	C-GCCAGGGT	CTTGCATACG	CTCTGGC-GC	GGTTAAAT
S. Typhi	GGCTGGTAAC	CATTC-ATAT	TCGCATTATT	GAACGATCAC	CGGCCAACCG	C-GCCAGGGT	CTTGCATACG	CTCTGGC-GC	GGTTAAAT
S. Paratyphi A	GGCTGGTAAC	CATTC-ATAT	TCGCATTATT	GAACAATCAT	CGGCCAACCG	C-GCCAGGGT	CTTGCATACG	CTCTGGC-GC	GGTTAAAT
Salmonella bondori	GGCTGGTAAC	CATTC-A-AT	TTGCATTATC	CAACCATCAC	CAGCCAGCCG	C-CCCAGGGT	CTTGCATACG	CTCTGGC-GC	CCTTADATC
E. coli 0157:H7	GGCTGGAAAT	CATTCAATAC	TCGCACTATC	GAAAGTTCAC	CAGCCAACCG	CAGCA-CGTT	CTTGCATACG	AGGTGCC-GC	GTTTTAC
Shigella flexneri	GGCTGGAAAT	CATTCAATAC	TCGCACTATC	GGAAGTTCAC	CAGCCAGCCG	CAGCA-CGTT	CTTGCATACG	ACGTGTCTGC	GGTCT-C
Citrobacter koseri	GGCTGGCAAC	CGTTCTATAT	TCGCACTATC	GAACGATCGC	CAGCTGGCCG	CAGTTGCGTT	CTTGCATACG	ACGCAACTGC	GGCAA

#### Figure 4.2 Predicted structure and conservation of STnc520

**A.** Software from www.nupack.org was used to predict the structure of the STnc520 sRNA. The colours of each base-pair indicate the probability that this pairing, or no pairing, occurs. The colour bar to the right of the predicted structure indicates the colour used for each probability score. STnc520 is predicted to form two stable stem-loop structures with a minimum free energy of -37.40 kcal/mol. Numbers correspond to the distance of that nucleotide from the TSS (+1) of the STnc520 gene. **B.** Multiple alignment of the STnc520 sequence from ten enteric bacteria. Numbers correspond to the distance of that nucleotide from the +1 site of the STnc520 gene and can be directly compared to the predicted structure shown in panel (**A**). Red indicates >90% sequence nucleotide identity, blue indicates >50% <90% nucleotide sequence identity between the 10 bacterial genomes.

#### 4.2.3 STnc520 shows a SPI1-like pattern of expression

As previously discussed (section 3.2.10.1), STnc520 shows a SPI1-like pattern of regulation and has a high correlation to the archetypical SPI1 gene, *prgH*, across a panel of 18 regulatory mutants (Pearson correlation coefficient: 0.87). Correlative analysis of *prgH* and STnc520 expression in wild-type across a range of 22 environmental conditions (Kröger et al., 2013) produces a Pearson correlation co-efficient of 0.96, suggesting that the promoter of STnc520 receives many of the same environmental inputs as SPI1 gene promoters. Figure 4.3 A is a screenshot from the online gene expression compendium, SalCom, (Kröger et al., 2013) showing the absolute expression of the STnc520 and *prgH* genes. While STnc520 is a more abundant transcript than *prgH* under all conditions, both genes are most highly induced at ESP, during anaerobic growth and after an aerobic shock. These conditions mimic aspects of the host intestinal environment, and stimulate the induction of invasion-associated genes.

Data presented in chapter 3 suggest that STnc520 has 5 regulatory inputs (2.5-fold or greater change in expression in 5 regulatory mutants). In addition to a dependence on HilD for expression (Figure 3.6 A), STnc520 expression is also reduced in the absence of BarA/SirA, FliZ and Fur, while STnc520 expression is increased in the absence of PhoP/Q (Figure 4.3 B and C). Expression of hilD is reduced in the other SPI1-associated regulatory mutants and increased in the  $\Delta phoP/Q$  mutant. It has previously been shown that the BarA/SirA TCS affects the transcription of hilD, while Fur and FliZ are likely to act at the level of the HilD protein, perhaps affecting activity. HilD auto-regulates expression of hilD, and mutations which affect HilD activity will also result in a reduction of transcription of the hilD gene (Ellermeier & Slauch, 2008, Ellermeier et al., 2005, Chubiz et al., 2010). PhoP represses SPI1 through HilA rather than HilD, but it has been suggested that PhoP may also activate expression of HilE, a repressor of HilD activity (Ellermeier & Slauch, 2007). It is, therefore, likely that the differential expression of STnc520 in the  $\Delta barA/sirA$ ,  $\Delta fur$ ,  $\Delta fliZ$  and  $\Delta phoP/Q$  mutant strains is the indirect result of differential hilD expression. Based on the high correlation of STnc520 expression and regulatory patterns with the SPI1 gene prgH, it was decided to investigate if HilD is the direct regulator of STnc520 transcription.

A.

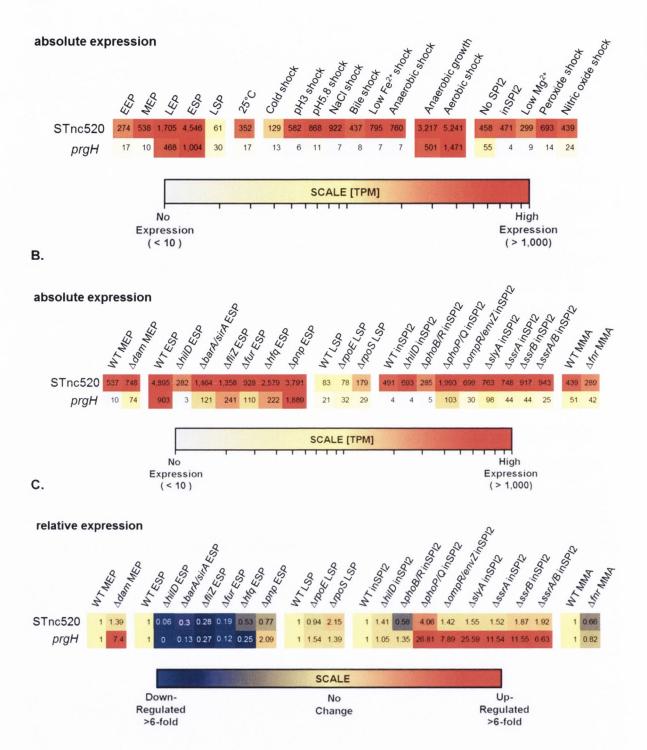


Figure 4.3 STnc520 has a SPI1-like pattern of expression and regulation

**A.** Sreenshot from the SalCom website (Kröger *et al.*, 2013) showing absolute expression (TPM values) of STnc520 and the archetypical SPI1 gene *prgH* under 22 environmental conditions. Expression of both genes is most highly induced at ESP, during anaerobic growth and subsequent aerobic shock. **B.** TPM values showing absolute and **C.** relative expression of the STnc520 gene in 18 regulatory mutants and their wild-type comparators. STnc520 is differentially expressed in mutants lacking the HilD, BarA/SirA, FliZ, Fur and PhoP/Q regulatory proteins.

## 4.2.4 Investigation of the role of HilD in regulating STnc520 expression by exogenous expression analysis

To investigate the role of HilD in regulating expression of STnc520, HilD was ectopically expressed from the arabinose-inducible P<sub>BAD</sub> promoter on the pBAD vector, in a variety of mutant backgrounds, as previously discussed (section 2.2.3). Preliminary studies found that STnc520 expression was sensitive to activation by basal levels of HilD produced due to the leaky nature of the ParaBAD promoter (Guzman et al., 1995). Glucose can be used to repress expression from the P<sub>BAD</sub> promoter by reducing levels of 3',5'-cyclic AMP, thus lowering expression from the CRP-activated promoter of the positive regulator of the araBAD operon, AraC (Guzman et al., 1995). Each strain was grown with D-glucose to an OD<sub>600</sub>1.0. The cultures were then split into two and HilD expression was induced in one, by addition of L-arabinose, while the second one was left un-induced. Both cultures were grown for a further 45 minutes before transcription was stopped and cells were harvested for RNA extraction, as previously described (Pfeiffer et al., 2007). Strains carrying empty pBAD control vectors were treated in the same way. RNA was extracted as previously described (section 2.6.1). Total RNA was probed for STnc520 expression by northern blot.

Over-expression of HilD in a wild-type background led to a higher accumulation of STnc520 compared to the wild-type strain carrying an empty vector. Induction of HilD expression in a  $\Delta hilD$  mutant restored STnc520 expression to wild-type levels. The rescuing of STnc520 expression in the  $\Delta hilD$  mutant was specific, as STnc520 expression was not rescued in the  $\Delta hilD$  mutant with an empty vector or the non-induced pBAD-hilD vector (Figure 4.4 A). However, exogenously-expressed HilD was unable to restore STnc520 expression in a mutant lacking the entire SPI1 locus (Figure 4.4 B), suggesting that another SPI1-encoded factor must be involved in the activation of STnc520.

HilD is part of a feed-forward regulatory loop involving RtsA, which is encoded outside of SPI1, and HilC which is encoded on SPI1 (Ellermeier et al., 2005). We, therefore, investigated if the presence of HilC might be important for full STnc520 expression. HilD was over-expressed in a single  $\Delta hilC$  mutant and in a double  $\Delta hilD\Delta hilC$  mutant. STnc520 expression was restored in both mutant backgrounds (Figure 4.4 C), indicating that HilC is not the SPI1-encoded factor that is necessary for STnc520 expression.

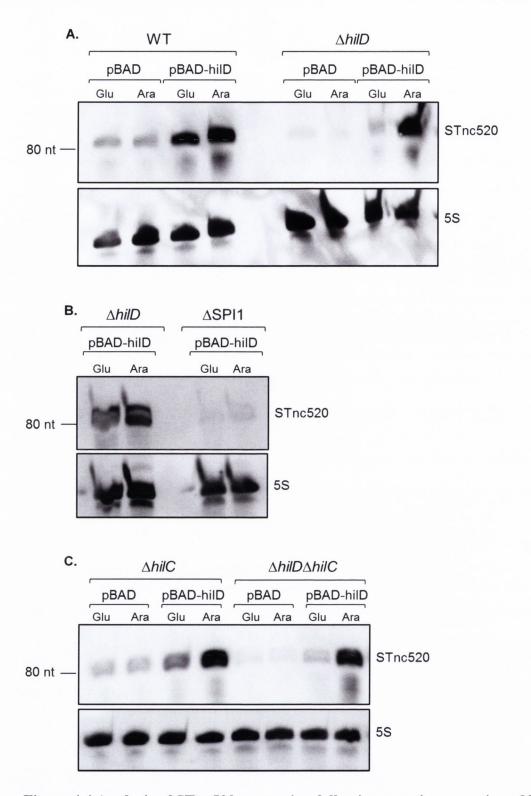


Figure 4.4 Analysis of STnc520 expression following ectopic expression of HilD

**A.** Northern blot showing that STnc520 is expressed in wild-type 4/74 and accumulates to a higher level when HilD is produced exogenously. STnc520 is not expressed in a  $\Delta hilD$  mutant, but expression is specifically restored when HilD expression is induced from the  $P_{BAD}$  promoter and is not restored in the absence of L-arabinose or upon L-arabinose induction in a  $\Delta hilD$  mutant carrying an empty pBAD vector. **B.** Ectopic expression of HilD does not restore STnc520 expression in a mutant lacking the entire SPI1 locus. **C.** Ectopic expression of HilD restores STnc520 expression in  $\Delta hilC$  and  $\Delta hilD\Delta hilC$  mutants. "Glu" refers to D-glucose which was used, at a final concentration of 0.2%, to repress the system and prevent basal levels of leaky HilD expression, "Ara" refers to L-arabinose which was used, at a final concentration of 0.2%, to induce the system, as previously described (section 2.2.3). All northern blots were probed for 5S ribosomal RNA as a loading control.

#### 4.2.5 SprB is the SPI1-encoded transcription factor necessary for STnc520 expression

HilD acts at the top of the SPI1 regulatory hierarchy, and the expression of a number of SPI1-encoded and SPI1-associated regulatory proteins would be affected in a ΔhilD mutant, which could explain the decrease in STnc520 expression that was not rescued by ectopic expression of HilD in a ΔSPI1 mutant. In addition to the five SPI1-encoded TFs (discussed in section 4.1.2), SPI1 is positively regulated by the RtsA/B TCS, as part of the feed-forward regulatory loop with HilD and HilC. SPI1 is also negatively regulated by the interaction of HilE with HilD (Baxter et al., 2003). Neither of these regulatory elements are encoded on SPI1. RNA was extracted from mutants lacking these SPI1-encoded and SPI1-associated regulators grown to ESP, and the RNA was probed for expression of STnc520 by northern blot to identify the SPI1-encoded or SPI1-associated factor involved in direct activation of STnc520 transcription. A ΔSTnc520 mutant was also included to demonstrate the specificity of the STnc520 riboprobe.

As previously shown, STnc520 expression is decreased in the absence of HilD. Expression of STnc520 is not affected in the absence of the HilC, RtsA/B, InvF or HilE proteins and STnc520 expression appears to be slightly higher in the absence of the HilA protein. STnc520 is not expressed in the \( \Delta sprB \) mutant, indicating that SprB is the additional SPI1-encoded TF responsible for activating the expression of STnc520 (Figure 4.5). RNA-seq data from this study shows that sprB is not expressed in the  $\triangle hilD$  mutant (Appendix II), and a recent investigation of HilD binding sites, using ChIP-seq, determined an association of HilD with the sprB promoter in S. Typhimurium (Petrone et The HilD dependency of sprB accounts for the decrease in STnc520 expression seen in the  $\triangle hilD$  mutant. However sprB expression has also been reported to be HilA-dependent (Saini & Rao, 2010), which does not reflect the slight increase in STnc520 expression in the  $\triangle hilA$  mutant, seen in Figure 4.5. However, unpublished RNA-seq data from our laboratory, which profiled the transcriptome of a  $\Delta hilA$  mutant grown to an OD<sub>600</sub>0.8 in Lennox broth, and compared to the wild-type strain grown under the same conditions, shows only a modest decrease in sprB expression in the absence of HilA, and no significant change in hilD expression. These data suggest that sprB transcription is not fully dependent upon HilA and that, in the absence of HilA, HilD could become the key regulator of sprB, and thus STnc520, expression. However, it cannot be ruled out that other non-SPI1-associated TFs could also be involved in regulating transcription of STnc520.

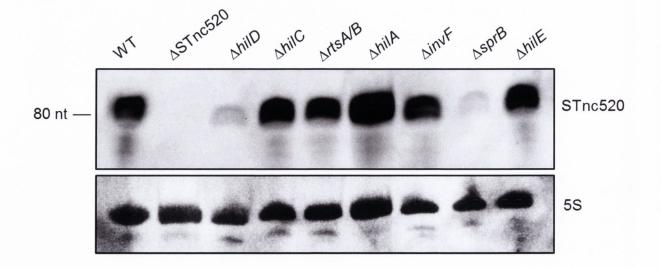


Figure 4.5 Expression of STnc520 in isogenic mutants of the key SPI1-encoded and SPI1-associated transcription factors

Northern blot showing STnc520 expression in all SPI1-encoded TFs and in the SPI1-associated TFs RtsA/B and HilE. A mutant lacking the STnc520 gene was also included to demonstrate probe specificity. STnc520 is not expressed in mutants lacking the HilD and SprB regulatory proteins, while STnc520 expression is slightly higher in a mutant lacking the HilA. The blot was probed for 5S ribosomal RNA as a loading control.

#### 4.2.6 Analysis of SprB binding to the STnc520 promoter

#### 4.2.6.1 Construction of a 4/74 sprB-FLAG strain

Unpublished experiments to identify SprB binding sites within the *S.* Typhimurium 14028 chromosome, using ChIP-seq with an *sprB*-FLAG strain, identified an SprB binding site downstream of *pliC*, in a region that corresponds to the 4/74 STnc520 promoter (Professor Joseph T. Wade pers. comm.). There is 100% nucleotide sequence identity in this region between *S.* Typhimurium 14028 and 4/74. For the global ChIP study, a C-terminal FLAG tag was added to the *sprB* gene in its chromosomal location in 14028, using the "Flexible Recombineering Using Integration of *thyA*" (FRUIT) recombineering method (Stringer *et al.*, 2012). This method relies upon the *thyA* gene for the production of thymine as a selectable and counter-selectable marker for gene deletion, tagging, point mutations or promoter swapping, and results in genetically heritable and scarless chromosomal mutations. Scarless mutagenesis is important in gene tagging to reduce the chance that the tag or selectable marker will cause decreased activity of the gene product, and so permit the functionality of the protein to be investigated.

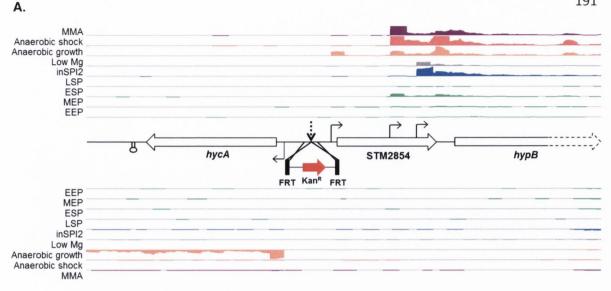
To investigate direct binding of SprB to the STnc520 promoter in 4/74 it was necessary to move the FRUIT-mediated FLAG tagged sprB gene from 14028 into 4/74. Direct P22 phage transduction was not possible, as no co-located antibiotic resistance cassette was available for use as a selectable marker. In addition, preliminary data indicated that the FRUIT method may not work in the 4/74 genetic background (J.T. Wade, pers. comm.). Consequently, a Kan<sup>R</sup> cassette was introduced in a transcriptionally silent region of the chromosome of 14028 sprB-FLAG, without removing any DNA, via λ-Red recombination, using oligonucleotides listed in Appendix IX. The Kan<sup>R</sup> cassette was introduced in the IGR between the divergently transcribed hycA and STM2854 genes, a region which is transcriptionally silent in the wild-type strain under 22 environmental conditions (Figure 4.6 A) (Kröger et al., 2013). The packaging capability of the P22 HT 105/1 int-201 phage is approximately 44 kb (Casjens & Hayden, 1988). The Kan<sup>R</sup> cassette was integrated approximately 12 kb from start of the tagged sprB gene (Figure 4.6B) and was used to co-transduce the tagged sprB gene into 4/74 by P22 phage transduction, generating strain JH3778 (Table 2.2). 4/74 transductants were screened for the presence of the FLAG tag by colony PCR, using sprB\_check oligonucleotides (Appendix IX). The functionality of the tagged SprB protein was assessed in terms of STnc520 transcription. RNA from wild-type 4/74 and 4/74 *sprB*-FLAG, grown to ESP, was probed for STnc520 expression by northern blot. STnc520 expression was not affected in the 4/74 *sprB*-FLAG strain (Figure 4.6 C).

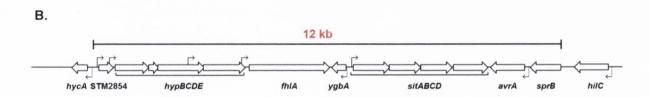
#### 4.2.6.2 SprB binds the STnc520 promoter in vivo

Chromatin Immunopreciptation followed by quantitative real-time PCR (ChIP-qPCR) was carried out to investigate if SprB binds to the STnc520 promoter region in 4/74. The *sprB*-FLAG strain was grown to ESP and the ChIP assay was carried out, as previously described (section 2.8) and qPCR was carried out, as previously described (section 2.6.4). The *hemX* gene was used as a negative control region.

The amplified ChIP DNA was normalised to the amplified Input DNA, which had been extracted prior to immunoprecipitation (IP/Input). Figure 4.7 A shows the normalised ChIP-qPCR data from two independent biological replicate experiments and demonstrates that there is strong enrichment (approximately 14-fold) of the STnc520 promoter region in the experimental ChIP DNA, compared to the background "mock" ChIP DNA. The negative control gene, *hemX*, displayed little enrichment in the experimental ChIP DNA sample, compared to the mock ChIP DNA sample. Following subtraction of the mock ChIP DNA, the STnc520 promoter DNA was approximately 9-fold enriched for SprB binding, compared to the negative control region, *hemX* (see formula in section 2.8.5) (Figure 4.7 B). These data show that SprB specifically binds within the STnc520 promoter region and is a direct activator of STnc520 transcription.







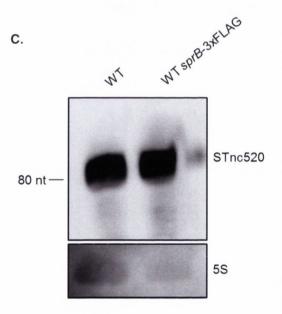
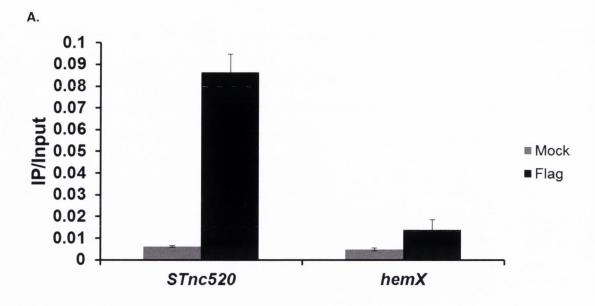


Figure 4.6 Construction and functional activity of 4/74 sprB-FLAG

A. Screenshot from Integrated Genome Browser (IGB) showing RNA-seq data from wild-type 4/74 grown under 9 environmental conditions (this study and (Kröger et al., 2013)) demonstrating that the IGR between hycA and STM2854 is transcriptionally silent. The dotted vertical arrow marks the insertion site of a kanamycin resistance cassette by λ-Red-mediated recombination (Datsenko & Wanner, 2000). Horizontal arrows indicate the direction of transcription of the indicated genes and dotted lines at the end of these arrows indicates that the genes are not shown to scale. B. Genetic map showing the approximately 12 kb region between the integration site of the Kan<sup>R</sup> cassette and the sprB gene. Symbols are as described in panel (A). C. Northern blot showing no differential STnc520 expression in wild-type 4/74 and 4/74 sprB-FLAG strains.



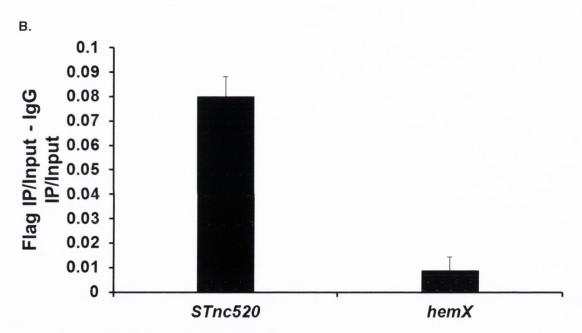


Figure 4.7 SprB directly binds to the STnc520 promoter region

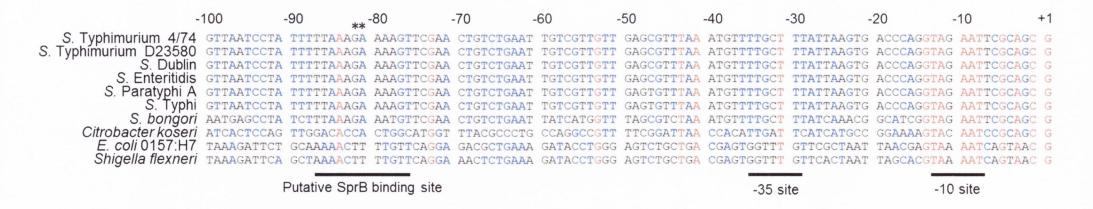
**A.** Chromatin immunoprecipitation followed by qPCR was used to investigate SprB binding to the STnc520 promoter *in vivo*. A "mock" immunoprecipitate using species-specific IgG was used to control for background levels of binding. Experimental and mock ChIP DNA was normalised to the starting amount of DNA (IP/Input). The STnc520 promoter was enriched in the experimental ChIP DNA, compared to the mock ChIP DNA. There was little enrichment of the *hemX* gene in the experimental ChIP DNA, compared to the mock ChIP DNA. **B.** The background (mock IgG) was subtracted from the experimental ChIP DNA. The STnc520 promoter was approximately 9-fold enriched for SprB binding, compared to the *hemX* negative control DNA. These data indicate that SprB binds the STnc520 promoter and directly regulates STnc520 transcription. Error bars are based on the standard deviation generated from 2 independent biological replicates.

## 4.2.7 Conservation of the STnc520 promoter region

SprB binds approximately 80 bp upstream of the putative STnc520 start site in 14028 (J.T. Wade pers. comm.). Based on motif analysis from the previously mentioned, unpublished global SprB ChIP-seq study, the most important nucleotides for SprB binding in the STnc520 promoter region are the G and A nucleotides at positions -81 and -82, relative to the STnc520 TSS (marked \* in Figure 4.8) (J.T. Wade pers. comm.).

The presence of conserved -10 sites upstream of the STnc520 TSS in the ten enteric bacterial species, shown in Figure 4.8, suggests that the STnc520 gene, or a variant of the gene, is transcribed in all of these species. However the lack of conservation of the -35 sites, particularly outside of the *Salmonella* genus, suggests that an alternative  $\sigma$ -factor may be required to activate transcription. The entire sequence of the putative SprB binding site is well conserved in the *Salmonella* genus and some nucleotides are highly conserved in other species. However the important GA nucleotides are not present in the species that lack an SprB orthologue, namely *Citrobacter koseri*, *E. coli* 0157:H7, *Shigella flexneri*.

As previously discussed in section 1.2.1, horizontal regulatory transfer involves the acquisition of non-homologous regulatory regions of DNA (Oren et al., 2014). The acquisition of alternative promoter regions can result in a re-wiring of the regulation of the downstream gene, through altering promoter architectures or TF binding sites. Horizontally-acquired promoter regions may provide the recipient bacteria with a fitness advantage in certain niches, even over closely related bacteria, and thus, HRT contributes to bacterial diversification (Oren et al., 2014). The re-wiring of gene regulation via adaptation of promoter sequences has been previously described for the ancestral Salmonella gene srfN. The promoter region of srfN contains an SsrB binding site in S. enterica that is not present in S. bongori. In addition, the function of SrfN differs from an unknown role in S. bongori to a role in intracellular fitness in S. enterica (Osborne et al., 2009). The SprB-specific sequences within the STnc520 promoter in the Salmonella genus may represent an example of HRT, and could present the interesting possibility that adaptive evolution of the STnc520 promoter region allowed co-option of the STnc520 sRNA function by the SPI1-encoded SprB, and that this event may have contributed to host cell invasion capabilities of the Salmonella genus when it diverged from its last common ancestor with E. coli.



# Figure 4.8 Multiple alignment of the STnc520 promoter sequence

Multiple alignment of the 100 bp upstream of the STnc520 gene orthologue in ten enteric bacteria. The TSS (+1), the -10 hexamer and the -35 hexamer are highlighted. SprB binds approximately 80 bp upstream of the STnc520 start site, at a site that is well conserved in the *Salmonella* genus and is less well conserved in *Citrobacter koseri*, *E. coli* 0157:H7 and *Shigella flexneri*, that do not contain an SprB orthologue. The indicated (\*) A and G nucleotides were found to be most important for SprB binding (J.T. Wade pers. comm.). Red indicates >90% sequence identity, blue indicates >50% <90% sequence identity, black indicates <50% sequence identity between the ten bacterial genomes.

#### 4.2.8 The role of STnc520 in virulence

The SPI1-like pattern of STnc520 expression and direct regulation of STnc520 transcription by a SPI1-encoded TF suggests that STnc520 could be involved in host epithelial cell invasion or in bacterial survival of the harsh conditions encountered in the host intestine. As previously discussed in section 3.2.12, a recent global mutagenesis study, which developed a technique known as Transposon-directed insertion-site sequencing (TraDIS), used next generation sequencing to sequence insertion sites of pools of 4/74 transposon mutants, following oral infection of chicken, pig and calf animal models, in comparison to an input inoculum grown in LB. The ratio of input to output reads for a particular insertion site was calculated to determine the fitness score of individual mutants. Negative fitness scores indicate an attenuation of the strain as a result of the transposon insertion (Chaudhuri et al., 2013). Two transposons were inserted into the STnc520 gene, and both insertions resulted in significant attenuation of the mutant strain after host colonisation. There was an approximately 5-fold decrease in fitness of both STnc520 transposon mutants in the chicken model. No output reads were received for the first insertion, which maps to the 3' end of the STnc520 gene, following pig and calf infection and this mutant was assigned an arbitrary negative fitness score of -15. The second insertion site caused an approximately 7-fold attenuation in the pig model. Again there were no reads obtained for the second insertion following infection of the calf model and the arbitrary negative fitness score of -15 was assigned (Chaudhuri et al., 2013). These data indicate that STnc520 plays a significant role in infection of the chicken, pig and especially of the calf via the oral route.

We investigated the secreted protein profile of the  $\Delta STnc520$  mutant compared to wild-type, grown to ESP, as this provides a measure of the functionality of the SPI1 secretion apparatus (Figure 4.9). There was no discernible difference in the secreted proteins between wild-type and  $\Delta STnc520$  mutant strains, indicating that STnc520 does not affect the SPI1 secretion apparatus or SPI1-secreted effector proteins during *in vitro* growth.

Thus, while STnc520 plays an important role in host colonisation and the infection process, STnc520 does not appear to do so by disrupting the function of the SPI1 TTSS.

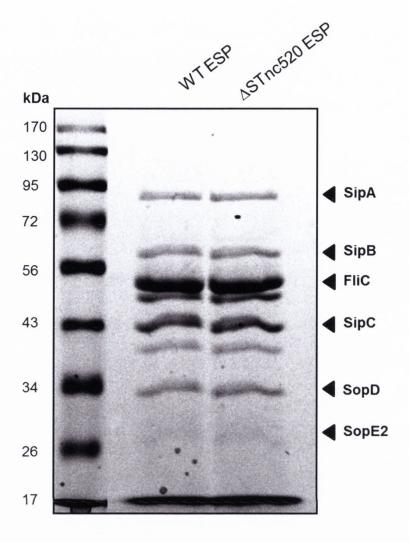


Figure 4.9 Investigation of the role of STnc520 on SPI1 secretion

Analysis of the secreted protein profile, by SDS-PAGE and Coomassie staining, of wild-type and ΔSTnc520, grown to ESP. Secreted protein fractions were prepared from culture supernatants, as previously described (section 2.7.2). The bands of known abundant effector proteins (SipA, SipB, SipC, SopD and SopE2) and of flagellin (FliC), are indicated according to (Raffatellu *et al.*, 2005). There is no change in secretion of the indicated SPI1 effector proteins or of FliC.

## 4.2.9 Investigation of mRNA targets of STnc520 using microarray-based technology

In order to investigate the putative role of STnc520 in S. Typhimurium virulence and within the SprB regulon, the identification of STnc520 target genes was required. Transcriptomic-based methods to identify the mRNA targets of sRNAs take advantage of the fact that many sRNA-mRNA interactions result in the rapid degradation of the mRNA Typically, transcriptomic-based target-hunting target (Vogel & Wagner, 2007). experiments involve a short or pulse over-expression expression of the sRNA, followed by scoring of the changes in expression of genes by microarray. The short pulse of the sRNA is designed to show changes in gene expression of direct sRNA targets, as opposed to those affected by downstream regulatory effects (Papenfort et al., 2006). Unpublished target-hunting microarray-based data (Händler, 2014) was available for STnc520. Briefly, STnc520 was pulse over-expressed for 10 minutes from the arabinose inducible P<sub>BAD</sub> promoter in a ΔSTnc520 mutant background at ESP. The ΔSTnc520 mutant background was used to remove regulatory effects which could be mediated by chromosomal STnc520, and only allow direct effects following the short over-expression of STnc520 to be detected. As STnc520 is most highly expressed at ESP it was assumed that STnc520 is most functional at this growth condition. Transcriptomic changes following over-expression of STnc520 were compared to a  $\Delta$ STnc520 mutant carrying pKP-8-35, a pBAD control vector which expresses a nonsense RNA when induced with arabinose (Papenfort et al., 2006). Table 4.1 lists the genes which showed significant (>2-fold) and reproducible (p < 0.05 in 2 replicates) differential expression, when STnc520 was transiently over-expressed in comparison to the over-expression of a nonsense RNA.

Table 4.1 also shows the fold change of the differentially expressed genes in the  $\Delta hilD$  mutant, grown to ESP, (determined by RNA-seq) because STnc520 expression is 17-fold reduced in the absence of HilD (Figure 4.3 B and C). sRNAs may act as the missing link between TFs and genes within the regulon of that TF, that are not directly regulated by the TF, as has been shown in the case of RyhB and the Fur regulon (Masse et al., 2005). We speculate that sRNAs, such as STnc520, could be the missing regulatory link between genes that are differentially expressed in the absence of HilD but are not directly regulated by HilD. The reduction in STnc520 expression in the  $\Delta hilD$  mutant background may have a similar effect as a  $\Delta$ STnc520 mutation on STnc520 target genes. Comparison of genes which are differentially expressed in the absence of HilD with

genes that are differentially expressed upon pulse over-expression of STnc520 may provide additional evidence that a regulatory interaction occurs. Expression of purM, purN and cvpA is down-regulated by over-expression of STnc520 and is also up-regulated in the  $\Delta hilD$  mutant. HilD has never been reported to act as a transcriptional repressor, thus the up-regulation of these genes in the  $\Delta hilD$  mutant is likely to be an indirect effect and could be mediated by STnc520. However the other genes which are not inversely expressed in the absence of HilD, compared to over-expression of STnc520, cannot be ruled out as being STnc520 targets under some environmental condition. In addition, functional redundancy between STnc520 and other regulators may also mask the effects of the reduction in STnc520 expression in the hilD mutant background.

Table 4.1 Genes showing significant (>2-fold) and reproducible (p < 0.05) differential expression following pulse over-expression of STnc520 and microarray

Gene	Fold Change (microarray <sup>a</sup> ) pSTnc520/pKP-8-35 <sup>b</sup>	Fold Change (RNA-seq) ΔhilD/WT <sup>c</sup>	Description		
ompW	-3.9	-1.6	Hypothetical outer membrane protein		
purM	-3.5	2.1	Phosphoribosylformylglycinamidine cyclo-ligase		
yfiD	-3.5	-1.8	Conserved hypothetical protein		
orfX	-3.4	-4.1	Conserved hypothetical protein		
napG	-2.6	-7.4	Ferredoxin-type protein NapG		
purN	-2.5	2.1	Phosphoribosylglycinamide formyltransferase		
cvpA	-2.3	1.6	Colicin V production protein (DedE protein)		
secG	2.2	-1.03	Protein-export membrane protein		
hslT	3.6	-1.02	Heat shock protein A		

<sup>&</sup>lt;sup>a</sup> Data from (Händler, 2014)

 $<sup>^{\</sup>rm b}$  Fold change in expression following over-expression of STnc520, compared to a nonsense RNA from the  $P_{\rm BAD}$  promoter

<sup>&</sup>lt;sup>c</sup> Fold change in expression in  $\Delta hilD$  mutant, compared to wild-type

#### 4.2.10 Investigation of the interaction between STnc520 and putative target mRNAs

The validity of six of the putative gene targets of STnc520 was investigated using the GFP-based two-plasmid target validation system, as previously described (section 2.9) (Urban & Vogel, 2007, Corcoran et al., 2012). This method of validation is based on the canonical model of sRNA-mediated regulation, that involves the sRNA binding its target mRNA close to the ribosome binding site or start codon, and affecting translation and/or degradation of the target mRNA. In the optimised two-plasmid validation system, described by Corcoran et al, the 5' UTR and 5' of the target gene are fused to Superfolder GFP under the control of a constitutively expressed P<sub>LtetO</sub> promoter (pXG10-sf vector). The entire sRNA is cloned under the control of a constitutively expressed P<sub>LlacO</sub> promoter on a second plasmid (pP<sub>L</sub> vector). The target fragment and sRNA are then co-expressed in a heterologous system (E. coli TOP10 cells). The action of the sRNA on expression of the cognate target mRNA interferes with GFP accumulation in the cell, and the fluorescent output can be measured. Fluorescent output, following co-expression of the putative target mRNA with the sRNA, is compared to fluorescent output, following co-expression of the target mRNA with the control plasmid, JV300, comprising a nonsense RNA also expressed from the P<sub>LlacO</sub> promoter (Sittka et al., 2007).

The ompW, purM, orfX, secG and hslT putative gene targets were investigated using the two-plasmid target validation system. napF, the first gene in operon that contains napG was also investigated in this way as napF expression is also down-regulated in the  $\Delta hilD$  mutant.  $E.\ coli$  TOP10 cells containing each of the pXG10 fusions (detailed in Appendix VII) (Händler, 2014) and either the pP<sub>L</sub>-STnc520 or JV300 vectors were grown overnight in Lennox broth. Whole cell lysates were generated as previously described (section 2.7.1) and fluorescent output was measured by western blot (section 2.7.5). DnaK was used as a loading control.

Constitutive expression of STnc520 did not cause any change in the level of GFP expression for any of the tested constructs, compared to the constitutive expression of a nonsense RNA (Figure 4.10). This finding could either indicate that STnc520 is not a direct regulator of these genes or that the region necessary for target regulation by STnc520 was located outside of the region cloned into the pXG10 plasmid. It was decided to refine the target hunting experiment to further the identification of direct STnc520 targets.

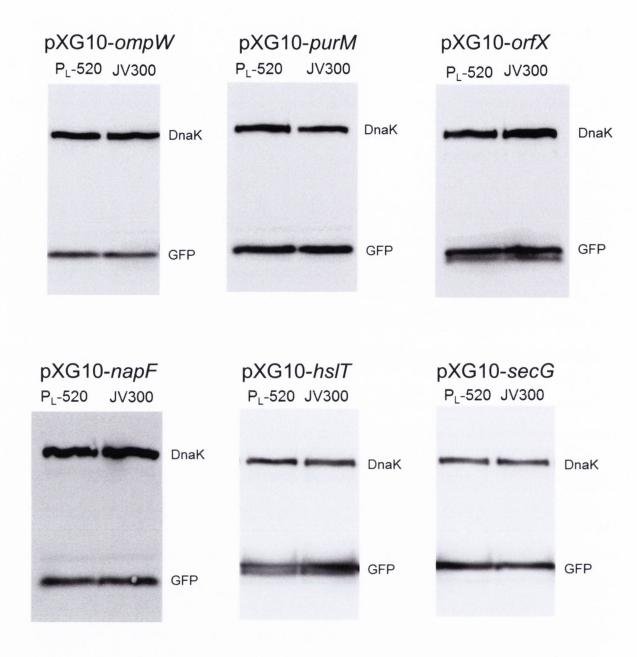


Figure 4.10 Use of the two-plasmid system to investigate putative mRNA targets of STnc520

Western blots showing accumulation of GFP in the indicated translational fusion reporter plasmids (pXG10-ompW/purM/orfX/napF/hslT/secG) when constitutively co-expressed with STnc520 or a nonsense RNA on a second plasmid. GFP accumulation is not affected by STnc520 in any of these reporter plasmid strains, suggesting these genes are not direct targets of STnc520, or that STnc520 targets the mRNA in a non-canonical fashion i.e. by binding outside the RBS and 5' region of the target mRNA.

#### 4.2.11 Investigation of mRNA targets of STnc520 using RNA-seq-based technology

Methods that are currently used to search for sRNA targets tend to generate a large number of false positives (Vogel & Wagner, 2007). This could be due to titration of Hfq as was shown following over-expression of ArcZ (Papenfort et al., 2009) or, if the sRNA regulates expression of a regulatory protein, many secondary regulatory effects may be detected (Vogel & Wagner, 2007). sRNA-mediated gene regulation occurs more rapidly than regulation by a TF as the sRNA does not require translation before becoming active. The 10 minute pulse expression of the sRNA used in the current microarray-based target-hunting experiments is designed to take advantage of rapid sRNA-mediated regulation, and to minimise secondary effects (Papenfort et al., 2006, Händler, 2014). However, the unsuccessful attempt to validate many of the predicted STnc520 targets (Figure 4.10) suggests that a large number of false positives are still being detected by this method. The relatively insensitive microarray-based method for monitoring the transcriptome may also fail to detect subtle changes in gene expression that could result from redundancy of sRNA function. We, therefore, repeated the sRNA target-hunt using a shorter pulse time of 5 minutes, and transcriptomic changes were measured with RNA-seq rather than microarray.

Wild-type and ΔSTnc520 carrying an empty pBAD vector and ΔSTnc520 pBAD-STnc520 were grown to ESP, and L-arabinose was added to a final concentration of 0.2%. Preliminary experiments showed that addition of glucose was not necessary, as there was no leaky expression of STnc520 prior to induction, and that a 5 minute induction was sufficient to over-express STnc520 (data not shown). STnc520 expression in samples used for the RNA-seq based target-hunting experiment is shown by northern blot in Figure 4.11 A. The cDNA library preparation and sequencing on the Illumina MiSeq platform was carried out as described in sections 2.6.7 and 2.6.8. Small transcripts (<200 nt) were lost during sample preparation, however this was unimportant for this sequencing run as the mRNA rather than sRNA targets of STnc520 were under investigation. The pipelines for mapping and calculation of TPM values were performed, as previously described (section 2.6.9).

We wished to identify genes that were most likely to be direct targets of STnc520, based on the pattern of expression of these genes in both the  $\Delta$ STnc520 mutant and following pulse over-expression of STnc520. Genes which were differentially expressed in the

absence of STnc520, and showed an inverse pattern of expression, or restoration to wild-type levels of expression, upon over-expression of STnc520 were considered to be likely candidates for direct regulation by STnc520, according to the criteria detailed in Table 4.2. As previously discussed (section 4.2.9), the reduction in expression in STnc520 in the absence of HilD may produce similar effects to a  $\Delta$ STnc520 mutant. However, potential functional redundancy of STnc520 and the control of expression of other regulatory proteins by HilD complicates the interpretation of the *hilD* mutant data with respect to the STnc520 target-hunting RNA-seq-based data. The *hilD* mutant data is, therefore, supplied for reference purposes only.

Table 4.2 Interpretation of RNA-seq-based STnc520 target-hunting data

Expression <sup>a</sup> in ΔSTnc520/WT <sup>b</sup>	Expression in ΔSTnc520 pSTnc520°/ΔSTnc520	Expression in $\Delta hilD/WT^d$	Hypothesis
+		+	Genes are negatively
•			regulated by STnc520
	+	-	Genes are positively
-			regulated by STnc520

<sup>&</sup>lt;sup>a</sup> +/- represents an up-regulation or down-regulation in gene expression of indicated strains, compared to comparator strain

Seventy-five genes were differentially expressed (>2-fold) in the  $\Delta STnc520$  mutant, compared to wild-type. Of the differentially regulated genes, 25 genes were up-regulated and 50 genes were down-regulated in the  $\Delta STnc520$  mutant. Twenty four genes were differentially expressed when STnc520 was pulse over-expressed in the  $\Delta STnc520$  mutant, compared to the  $\Delta STnc520$  strain carrying an empty pBAD vector (18 genes up-regulated and 6 genes down-regulated) (Figure 4.11 B and Appendix V). Only one gene, orfX, which had been previously identified as a putative mRNA target of STnc520 (Table 4.1) was also differentially expressed in this experiment. Both experiments

<sup>&</sup>lt;sup>b</sup> ΔSTnc520 and WT strains carrying empty pBAD vectors, and induced with 0.2% L-arabinose for 5 minutes

 $<sup>^{\</sup>rm c}$  STnc520 expressed from the  $P_{BAD}$  promoter following induction with 0.2% L-arabinose for 5 minutes

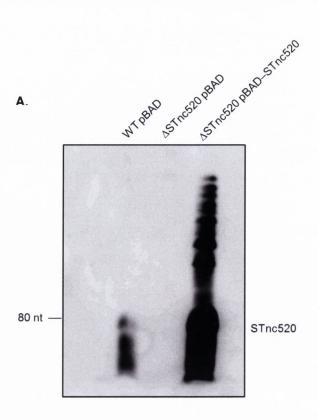
 $<sup>^{\</sup>mathrm{d}}$  hilD mutant data was not necessarily used for prediction of direct STnc520 targets

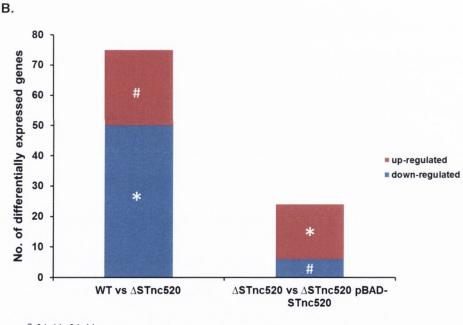
showed a decrease in orfX expression following pulse over-expression of STnc520, however orfX expression was also reduced >2-fold in the  $\Delta$ STnc520 mutant, compared to wild-type (Appendix V). As expression of orfX was differentially expressed, in microarray and RNA-seq-based target-hunting experiments, STnc520 is likely to be a regulator of the pathogenicity island-encoded orfX. However, the expression pattern of orfX does not reflect our hypothesis of direct regulation by STnc520, as presented in Table 4.2. Therefore, the regulatory interaction between STnc520 and orfX may be complex and indirect, as Figure 4.10 also suggests.

Of the genes which were up-regulated in the  $\Delta STnc520$  mutant, compared to wild-type, two genes (fdnH and fdoH) were also down-regulated by over-expressing STnc520, compared to  $\Delta STnc520$ . Of the genes which were down-regulated in the  $\Delta STnc520$  mutant, compared to wild-type, 9 genes were up-regulated by over-expressing STnc520, compared to  $\Delta STnc520$  (Figure 4.11 B). Six of the latter group of 9 genes are involved in sulphate metabolism and are encoded in 5 transcriptional units. The other 3 genes in this latter group are involved in copper homeostasis and are transcribed from 3 transcriptional units. These 11 genes fulfil the criteria of our hypothesis and are likely candidate genes for direct regulation by STnc520. These candidate target genes are detailed in Table 4.3.

Eight genes that were down-regulated in the  $\Delta STnc520$  mutant, compared to wild-type, were also down-regulated in the  $\Delta hilD$  mutant (Appendices II and V). Three of these genes include the SPI2-encoded ssaG, ssaJ and ssaI, as well as the gene encoding the alternative flagellin subunit fljB. However, of the genes which were down-regulated in both  $\Delta STnc520$  and  $\Delta hilD$ , only ssaG expression was restored close to wild-type levels upon over-expression of STnc520, suggesting ssaG may be the only gene from this group which is directly targeted by STnc520 (Table 4.3). There are no genes which were up-regulated in both the  $\Delta hilD$  and  $\Delta STnc520$  mutants. There are 5 genes that were down-regulated in the absence of HilD and were up-regulated upon over-expression of STnc520 (Appendices II and V), however the expression pattern of these genes does not reflect the criteria detailed in Table 4.2, and we speculate that putative regulation of these genes by STnc520 is not direct.

Table 4.3 lists the likely candidate gene targets of STnc520, based on the pattern of expression of these genes in the  $\Delta$ STnc520 mutant and STnc520 over-expression strain. Gene expression in the  $\Delta$ hilD mutant is included for reference only.





# fdoH, fdnH

Figure 4.11 Investigation of the mRNA targets of STnc520 using RNA-seq

**A.** Northern blot showing expression of STnc520 in samples used for RNA-seq-based STnc520 target-hunt. Wild-type and  $\Delta$ STnc520 mutant carrying the empty pBAD vector and a  $\Delta$ STnc520 mutant carrying pBAD-STnc520 were grown to ESP and expression from the P<sub>BAD</sub> promoter was induced with 0.2% L-arabinose for 5 minutes before transcription was stopped and RNA was extracted. **B.** Bar chart showing the numbers of differentially expressed genes in the  $\Delta$ STnc520 mutant compared to wild-type and the numbers of differentially expressed genes when STnc520 is over-expressed in a  $\Delta$ STnc520 mutant background compared to the  $\Delta$ STnc520 mutant carrying an empty pBAD vector. Genes which we speculate may be direct targets of STnc520 based on their patterns of expression are highlighted (\*/#).

<sup>\*</sup> golB, copA, cysW, cysP, cysC, cysN, cysl, cueP, sbp

Table 4.3 RNA-seq data showing differential expression of genes in  $\Delta STnc520$  or following pulse over-expression of  $STnc520^a$ 

Gene	WT pBAD	ΔSTnc520	ΔSTnc520	Fold Change in	Description		
Gene		pBAD pBAD-STnc520		$\Delta hilD$	•		
Genes up-regulate	ed in $\Delta STnc520$ ar						
	10	22	11	-2.2	formate dehydrogenase, nitrate-inducible, iron-sulphur		
fdnH					subunit		
	10	49	21	1.1	formate dehydrogenase-O beta subunit, iron-sulphur		
fdoH	19	49	21	1.1	subunit		
Genes down-regulated in ΔSTnc520 and restored by over-expressing STnc520							
STM0355/golB	26	12	26	-1.2	hypothetical copper chaperone		
copA	37	16	34	1.0	copper-transporting ATPase		
cysW	80	28	62	4.6	sulphate/thiosulfate transporter permease subunit		
cysP	67	26	76	5.6	thiosulphate-binding protein precursor		
cysC	44	15	33	5.1	Adenylylsulphate kinase		
cysN	29	12	33	3.2	Sulphate adenylyltransferase subunit 1		
cysI	23	10	24	5.0	sulphite reductase hemoprotein, NADPH dependent		
STM3650/cueP	118	50	107	1.2	anaerobic copper resistance protein		
sbp	38	11	31	2.3	periplasmic sulphate binding protein		
Genes down-regulated in $\Delta$ STnc520 and $\Delta$ hilD and restored by pulse over-expression of STnc520							
ssaG	207	99	147	-8.8	SPI2 Type III secretion system apparatus		

<sup>&</sup>lt;sup>a</sup> Heatmap indicates the absolute expression (in TPM) of each candidate target gene: blue indicates low expression, red indicates high expression

#### 4.2.12 Biocomputational prediction of STnc520 target binding sites

In silico methods of predicting sRNA binding sites have been developed to investigate the function of the vast amount of sRNAs which are being regularly identified (Vogel & Wagner, 2007). These bioinformatic approaches are based on calculating the energy required to form the short, and often imperfect, base-pairing interactions that define sRNA-mediated gene regulation. TargetRNA2 is a web server for the prediction of target mRNAs, based on the conservation of the sRNA across different species, the secondary structure of the sRNA and each candidate mRNA target and the hybridisation energy between the sRNA and each candidate mRNA target (Kery et al., 2014). The web-based IntaRNA programme predicts interactions between RNA molecules, taking target site accessibility and sRNA seed regions into account. IntaRNA calculates the combined energy score of opening the binding sites on both RNA molecules, to make the sites accessible for binding, and the energy of hybridisation (Busch et al., 2008). CopraRNA uses the IntaRNA algorithm to calculate target predictions, but CopraRNA also combines target prediction with phylogenetic information for homologous sRNAs from distinct organisms, to predict conserved regulatory interactions (Wright et al., 2013). TargetRNA2, IntaRNA and CopraRNA have been used to successfully predict a number of experimentally validated interactions between sRNAs and their target mRNAs (Kery et al., 2014, Wright et al., 2014), and are particularly useful to predict canonical binding of sRNAs around the Shine-Dalgarno site or start codon of the target mRNA.

TargetRNA2 and IntaRNA were used to perform whole genome screens to predict STnc520 binding sites within the 200nt upstream and 150 nt downstream of every coding gene in the Genbank database for *S.* Typhimurium 4/74. CopraRNA was used to perform whole genome screens to predict STnc520 binding sites within the 200 nt upstream and 150 nt downstream of every coding gene in the Genbank database for *S.* Typhimurium 4/74, *S.* Paratyphi B, *S.* Enteritidis and *S. bongori*. The highest scoring putative STnc520 binding sites predicted by each programme are listed in Appendix VI (TargetRNA2 (29 sites), IntaRNA (84 sites), CopraRNA (97 sites)). We speculate that targets which were predicted by one or more of the target prediction programmes are likely candidates to be direct STnc520 targets. Predicted targets that were differentially expressed in the absence of STnc520, and following over-expression of STnc520, were also speculated to be likely candidates for direct regulation by STnc520. Likely candidate genes, based on these criteria, are detailed in Table 4.4.

Table 4.4 Candidate STnc520 target genes based on bioinformatic analysis

Gene ID	Common Name	TargetRNA2 Energy (kcal/mol)	p-value	IntaRNA Energy (kcal/mol)	p-value	CopraRNA Energy (kcal/mol)	p-value	Description
SL3805	dnaA	-8.3	0.049	N/A	N/A	-10.4	0.00551	chromosomal replication initiator protein
SL3518	yhhA	-9	0.037	N/A	N/A	-9.1	0.02748	hypothetical protein
SL4086	secE	N/A <sup>a</sup>	N/A	-10.4	0.00783	-9.7	0.0083	preprotein translocase subunit SecE
SL0012	dnaK	N/A	N/A	-12.8	0.00019	-9.4	0.02955	molecular chaperone DnaK
SL1364	ydhO	N/A	N/A	-11.9	0.00093	-8.5	0.03121	putative cell wall-associated hydrolase
SL0651	gltL	N/A	N/A	-15	1E-07	N/A	N/A	glutamate/aspartate transport ATP-binding protein GltL
SL3170	glgS	N/A	N/A	N/A	N/A	-8.3	0.02108	glycogen synthesis protein GlgS
SL0350	golB	N/A	N/A	N/A	N/A	-11.9	0.00074	putative copper chaperone

<sup>&</sup>lt;sup>a</sup> N/A indicates that the gene was not predicted as a target by the particular target prediction programme

Two genes, *dnaA* and *yhhA*, were predicted to contain STnc520 binding sites by both TargetRNA2 and CopraRNA. Three genes, *secE*, *dnaK* and *ydhO*, were predicted to contain STnc520 interaction sites by IntaRNA and CopraRNA. There was no overlap in predicted target binding sites between TargetRNA2 and IntaRNA. The 5 genes that were predicted to contain STnc520 interaction sites by at least two target prediction programmes, and 96% of the genes listed in Appendix VI, were not differentially expressed in the absence of STnc520, or following over-expression of STnc520, in our RNA-seq-based target-hunt. However, it cannot be ruled out that STnc520 does target some of the non-differentially expressed genes, listed in Table 4.4 and Appendix VI, via the predicted interactions, but that the interactions only occur under different environmental conditions than ESP. It is also a possibility that STnc520 binding of these putative target mRNAs does not result in a change in stability of the mRNA, and thus, the regulatory interaction was not detected by our transcriptomic-based method.

Three genes predicted to contain STnc520 binding sites, by one of the target prediction programmes, were also differentially expressed in the absence of STnc520, and following over-expression of STnc520 (Appendix V). gltL (target predicted by IntaRNA) was differentially expressed (>2-fold) in the ΔSTnc520 mutant, compared to wild-type, in the RNA-seq-based target-hunt. However, gltL expression was not restored to wild-type levels upon pulse over-expression of STnc520 in the ΔSTnc520 background. putative interaction between STnc520 and gltL was the highest scoring interaction predicted by IntaRNA. It is possible that the 5 minute pulse over-expression of STnc520 was too short to mediate the STnc520-gltL interaction and to cause degradation of the gltL mRNA. Expression of glgS (target predicted by CopraRNA) is approximately 2-fold increased in the \Delta STnc520 mutant, compared to wild-type, but is also approximately 2-fold increased following over-expression of STnc520, compared to ΔSTnc520. The apparent opposing roles of STnc520 on glgS expression are difficult to reconcile with direct regulation by STnc520, but may suggest the involvement of additional regulatory factors. The third gene that contains a putative STnc520 interaction site (predicted by CopraRNA), and that was also differentially expressed in the RNA-seq-based target-hunting experiment, is golB. golB, which encodes a copper chaperone protein, is listed in Table 4.3 as a likely candidate target gene based on the pattern of golB expression in the RNA-seq-based target-hunting experiment, which fulfils the criteria detailed in Table 4.2. golB expression is reduced in the absence of STnc520 and is

restored to wild-type levels by over-expression of STnc520. These data suggest that STnc520 functions through stabilising the *golB* mRNA. The putative interaction between STnc520 and *golB* (Figure 4.12 A) is predicted to occur within the CDS of the upstream gene, which encodes the transcriptional regulator of genes involved in copper homeostasis, GolS. The predicted interaction site is located 165 nt upstream of the *golB* translation initiation codon, a region that suggests a non-canonical mechanism of regulation by STnc520. Further investigation of the putative interaction between STnc520 and *golB*, and the other likely candidate interaction sites listed in Table 4.4, is required to validate direct association of STnc520 at these sites.

With the exception of *golB*, none of the likely candidate genes, listed in Table 4.3, were predicted to contain a high scoring STnc520 interaction site within the 350 nt window investigated in the whole genome computational screens. It is possible, therefore, that the putative STnc520 interaction with these candidate target genes may also occur via a binding interaction in a non-canonical location.

# 4.2.13 Biocomputational prediction of STnc520 binding sites for candidate target genes

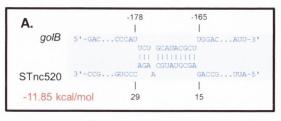
The IntaRNA programme can also be applied to non-whole genome screens and can be used to detect specific RNA-RNA interactions (Wright et al., 2014). IntaRNA was used to screen for STnc520 binding to target sites in the region upstream (>200 nt), and along the entire gene length of each of the candidate targets from Table 4.3. The highest scoring predictions are shown in Figure 4.12. The numbers above and below the gene sequence correspond to the nucleotide position, relative to the TSS of the sRNA and the first nucleotide of the translation initiation codon of the mRNA. The two highest scoring predicted interactions of STnc520 are with sbp and cysP, the genes which encode transport systems for the external sulphur sources, sulphate and thiosulphate respectively. The predicted binding sites are located over 220 nt upstream of the sbp start codon and over 370 nt downstream of the cysP start codon. Similar distances between translation initiation codons and the putative STnc520 recognition domains exist for cysC, cueP, fdnH, ssaG and copA. As discussed in section 1.3.1, sRNA-mediated regulatory interactions which occur outside of the translation initiation region are considered non-canonical interactions (Desnoyers et al., 2013). Increasing numbers of non-canonical sRNA-mediated regulatory interactions and mechanisms of action are being elucidated

(section 1.3.1) however, to our knowledge interaction sites at distances, such as those presented in Figure 4.12, from the translation initiation region have not, typically, been reported. The differential expression of the likely candidate genes in the RNA-seq-based target-hunt, coupled with the distances of the predicted STnc520 binding sites from the translation initiation regions of the candidate targets, suggests translation-independent regulation that affects target mRNA stability (Figure 1.4). Many of the putative regulatory interactions, presented in Figure 4.12, involve putative positive regulation by STnc520, which may occur in a translation-independent fashion through masking of ribonuclease binding sites, as has been demonstrated for RydC-mediated activation of the long isoform of *cfa* (Frohlich et al., 2013).

STnc520 is predicted to interact with sites downstream of the *cysJ* and *cysH* translation initiation codons, but neither of these genes were differentially regulated by STnc520 (Table 4.3). However, *cysJ* and *cysH* are encoded as part of a monocistronic operon with *cysI*, which is differentially regulated by deletion and over-expression of STnc520. *cysH* and *cysJ* also show the same pattern of expression as *cysI* in the ΔSTnc520 and pBAD-STnc520 strains, but *cysH* and *cysJ* have TPM values <10 and neither gene is considered to be expressed at ESP, based on the previously discussed criteria (section 2.6.9). However, the putative pairing between STnc520 and *cysJ* is the most canonical of the predicted STnc520 interactions, presented in Figure 4.12, as the putative interaction occurs within the 5'UTR of *cysJ*, the first gene in the *cysJIH* operon. The *cysJIH* operon is involved in the reduction of PAPS to sulphite and the reduction of sulphite to sulphide, as previously discussed (section 4.1.5). Because regulation of *cys* genes is controlled by sulphur levels, it is possible that interruption of just one stage of the sulphur reduction pathway will result in differential expression of the majority of genes in the pathway (Quan et al., 2002).

In all but two predicted interactions, nucleotides in the first stem loop of STnc520 (nt 1-44) are involved in target binding. As predicted in section 4.2.2, this loop region is an unstable structure and less energy is required to make the loop region accessible. The putative interactions of the sRNA with fdnH and ssaG occur in the loop of the  $\rho$ -independent terminator of STnc520 and are unlikely to be biologically relevant.

Figure 4.12 Predicted interactions between STnc520 and putative target mRNAs



```
-248
                                                -226
E.
        5'-GUU...GCGUC
                                                 AGUGC...UAA-3'
   sbp
                       CGUUCAA
                                  GCGA UGGAUGG
                                         11:1111
                       GCAAGUU
                                  CGCU
STnc520 3'-UAA...CACUA
                            AUCA
                                    UAU
                                                 AAUGG...CGG-5'
    -9.97 kcal/mol
                      35
                                                 9
```

```
G. +298 +305
| | | |
fdnH 5'-AAC...UGGUU ACUGU...GUA-3'
GUAUGC
| | | | | |
CAUACG

STnc520 3'-UAA...UCUCG UUCUG...CGG-5'
-7.2 kcal/mol 68 61
```

```
H. -165 -158
| | | | |
ssaG | 5'-CGU...GACCG | AGGUC...UGA-3'
GUAUGC
| | | | | |
CAUACG

STnc520 | 3'-UAA...UCUCG | UUCUG...CGG-5'
| | | |
-7.84 | kcal/mol | 68 | 61
```

## Figure 4.12 Predicted interactions between STnc520 and putative target mRNAs

**A.** Predicted interaction between STnc520 and the putative target mRNA *golB*, calculated by CopraRNA (section 2.10.3). **B-J.** Predicted interactions between STnc520 and the putative target mRNAs *cysP*, *cysJ*, *cysC*, *sbp*, *cueP*, *fdnH*, *ssaG*, *copA* and *cysH*, calculated by IntaRNA (section 2.10.3). Target mRNAs are the top sequence shown in each interacting pair. The numbers correspond to the distance from the first nucleotide (+1) of the translation initiation codon of the indicated protein-coding gene. The STnc520 sequence is shown at the bottom of the interacting pair and the numbers correspond to the distance from the TSS (+1) of STnc520. The combined free energy calculation of making the RNA interaction sites accessible and the energy of hybridisation is shown in red under each interacting pair.

## 4.3 Discussion

# 4.3.1 The transcriptional regulation of STnc520

STnc520 was identified as a SPI1-like sRNA based on its pattern of regulation (section 3.2.10) and STnc520 also shows a SPI1-like pattern of expression across a compendium of 22 physiologically relevant environmental conditions (Kröger et al., 2013). The 17-fold down-regulation of STnc520 expression in the absence of the primary SPI1 regulator, HilD, led us to investigate the relationship between HilD and STnc520. HilD is not directly involved in regulating expression of STnc520 but indirectly regulates STnc520, via SprB. The advantage of using a  $\Delta hilD$  mutant in our screen for the protein factors involved in transcriptional regulation of *S*.Typhimurium sRNAs is that HilD is at the top of the SPI1 regulatory cascade. Although HilD itself does not directly regulate expression of STnc520, the knowledge of the transcriptional regulators under the control of HilD allowed us to identify the factor that does directly regulate STnc520 expression. Thus, RNA-seq of just one mutant strain from the top of a regulatory hierarchy, rather than 6 downstream mutants, provided us with the information necessary to identify the direct regulator of this sRNA.

Very little is currently known about the SprB regulon. SprB is SPI1-encoded but does not play a role in the host intestinal epithelial cell invasion process or regulate expression of invasion-associated genes (Eichelberg et al., 1999). Although a slight repressive effect of SprB on *hilD* expression and binding of SprB to the *hilD* promoter region has been

reported after 12 hours of static growth in LB medium (Saini & Rao, 2010). The key regulatory effect that has been reported for SprB, however, is that SprB directly activates expression of the SPI4-encoded Type 1 secretion system that results in formation of a non-fimbrial adhesin, which is necessary for adhesion to the host intestinal cells prior to invasion (Saini & Rao, 2010, Gerlach et al., 2008). Direct binding of SprB to the STnc520 promoter was demonstrated using *in vivo* chromatin immunoprecipitation (ChIP) followed by qPCR. Thus, STnc520 can be added to the small number of genes which have been shown to be directly regulated by SprB and this may help to expand our knowledge about the role of SprB during the *Salmonella* infection process. As SprB has already been shown to control expression of genes on SPI4 and STnc520 is encoded on the pathogenicity island SPI11, it is tempting to suggest that SprB mediates cross-talk between SPI1 and other pathogenicity islands.

The STnc520 sequence is well conserved in sub-species of S. enterica and the sequence is partially conserved in S. bongori. The STnc520 sequence is also approximately 40% conserved in pathogenic E. coli, Shigella flexneri and Citrobacter koseri. The promoter region of the STnc520 gene is less well conserved outside of the Salmonella genus, in species that do not contain an SprB orthologue. This fact suggests that if the STnc520 sequence encodes a functional gene in other species, an alternative mechanism of transcriptional regulation of the gene is likely to be in use. Horizontal regulatory transfer may have played a role in the co-option of STnc520 function into the SprB regulon (Oren et al., 2014). The conservation of the STnc520 sequence within the Salmonella genus, as well as transcriptional activation of the STnc520 gene by the SPI1-encoded SprB, indicates that STnc520 could be important for Salmonella virulence. An important role for STnc520 during infection is further suggested by results from the recent transposon insertion-based global mutagenesis study (TraDIS) that identified 2 transposon insertions in the STnc520 gene associated with a fitness defect of the mutants during oral infection of food-producing animals (Chaudhuri et al., 2013). Like SprB however, STnc520 does not impact upon expression of SPI1 genes or secretion via the SPI1 apparatus at ESP. Based on the function of genes, in formate metabolism, sulphur metabolism and copper homeostasis, which are differentially expressed in the absence of STnc520, and following over-expression of STnc520, it seems plausible that STnc520 may be involved in maintaining metabolic homeostasis during infection.

#### 4.3.2 The role of STnc520 within the SprB regulon

Methods to validate direct interactions between sRNAs and their target genes have not kept up with the discovery of sRNAs which demonstrate non-canonical methods of gene regulation. Pulse over-expression of the sRNA followed by microarray, and more recently RNA-seq, has been used to identify sRNA-mediated transcriptomic changes. Transcriptomic-based methods of target identification are limited to the identification of interactions that result in changes in the stability of target mRNAs (Vogel & Wagner, 2007). Pulse over-expression is important to limit detection of indirect effects that may be caused by sRNAs which modulate expression of regulatory proteins. However, effects such as Hfq titration by strongly Hfq-bound sRNAs, can lead to dysregulation of other sRNAs and mRNAs upon sRNA over-expression (Papenfort et al., 2009). In addition, functional redundancy between sRNAs can lead to a masking effect, whereby deletion of one sRNA does not result in differential expression of a target, and deletion of both sRNA genes is necessary to produce an effect, as has been demonstrated for CsrB and CsrC-mediated control of invasion gene expression (Fortune et al., 2006). validation of a direct interaction between the sRNA and putative target mRNA is important to distinguish true targets from false positive and false negative targets. Putative sRNA-mediated regulatory interactions, identified by global transcriptomic analysis, can be validated using transcriptional fusions or qPCR, to measure target gene levels, or western blotting, to measure target protein abundance, in sRNA mutants or following over-expression of the sRNA. However, these methods still do not distinguish between direct and indirect regulatory mechanisms.

The two-plasmid validation system was developed as a method of confirming direct interactions between *trans*-acting sRNAs and target mRNAs (Urban & Vogel, 2007). This method involves constitutive co-expression of the sRNA of interest, from one plasmid, with a target region from the putative target mRNA, translationally fused to GFP on a second plasmid, in an exogenous system. GFP output can then be measured to assess if the sRNA can directly base-pair and affect translation of the mRNA. This method has typically been used to demonstrate sRNA binding within the target 5'UTR and repressing or activating target translation, often leading to ribonuclease-mediated target transcript degradation. To avoid affecting the correct folding or solubility of the GFP fusion protein, the minimum target sequence must be fused to the *gfp* gene. Additionally, full length putative target genes are not used in pXG10 vectors to avoid

including signal sequences for protein export. The lack of signal sequence ensures the maintenance of both plasmids together in the cytoplasm, allowing the putative interaction to occur (Urban & Vogel, 2007). The two-plasmid validation system can, therefore, not be used as effectively to confirm sRNA binding sites far upstream of the target RBS, within the target CDS or within the target 3'UTR. There have been a number of reported examples of sRNAs binding in these non-canonical positions, as previously discussed (section 1.3.1). Non-canonical interactions have typically been identified with *in silico* methods of predicting binding sites. Structure probing experiments and mutational analysis experiments are subsequently used to confirm the contribution of predicted binding sites to sRNA-mediated regulation (Corcoran et al., 2012, Pfeiffer et al., 2009). The studies by Corcoran *et al* and Pfeiffer *et al* used the two-plasmid validation system to confirm CDS binding of MicF and MicC, respectively, but the efficiency of the two-plasmid system decreases with increasing target sequence length.

STnc520 was first identified by co-immunoprecipitation with Hfq (Sittka et al., 2008). However, subsequent analysis has not classified STnc520 as a strongly Hfq-associated sRNA (Kröger et al., 2013). STnc520 does not require Hfq for stability, although the steady state levels of STnc520 decrease in a  $\Delta hfq$  mutant (Figure 4.1). We, therefore, speculate that STnc520 transiently binds Hfq and could have a regulatory role, binding target mRNAs in trans and likely requiring Hfq for activity. Two transcriptomic experiments were used to identify the mRNA targets of STnc520, to establish the role of STnc520 within the SprB regulon. The first experiment involved the pulse over-expression of STnc520 for 10 minutes, followed by microarray analysis of the changes in gene expression. Investigation of 5 out of the 9 most significantly and reproducibly differentially expressed genes was performed using the two-plasmid validation system. None of these putative targets were directly bound by STnc520 in the 5'UTR. A recent large scale study of the targets of S. Typhimurium sRNAs found that only 10% of the putative targets identified by microarray-based transcriptomic analysis, following pulse-expression of an sRNA, were validated using the two-plasmid validation system (Händler, 2014). The low success rate of target validation may reflect false positives arising from the microarray-based system, or may reflect the previously mentioned limitations of the two-plasmid validation system. A refined target-hunting approach was used instead, which involved a shorter pulse-expression of STnc520 for 5 minutes in a ΔSTnc520 background, followed by RNA-seq-based transcriptomic analysis,

compared to the  $\Delta STnc520$  mutant carrying an empty pBAD vector. The transcriptome of  $\Delta STnc520$  pBAD was also compared to wild-type pBAD, following a 5 minute induction of the  $P_{BAD}$  promoter.

Three groups of genes were considered promising candidates for direct regulation by STnc520, based on their patterns of expression in the absence of STnc520 and following over-expression of STnc520. The genes encoding the subunits of the oxygen and nitrate-inducible formate dehydrogenases were down-regulated by STnc520, while genes encoding the proteins involved in reduction of sulphur to cysteine and in copper homeostasis were up-regulated by STnc520. Finally, the gene encoding the major subunit of the SPI2 TTSS apparatus, ssaG, was down-regulated in the absence of STnc520 and ssaG expression was partially restored to wild-type levels by over-expression of STnc520 (Table 4.3).

Three computational screens of the 5'UTRs and 5' end of all genes in 4/74 did not predict canonical STnc520 binding sites that accounted for STnc520 regulation of these genes (Table 4.4). Therefore, the two-plasmid validation system was not used to attempt to validate direct interactions between STnc520 and the candidate targets. One high scoring STnc520 binding site was predicted far upstream of the candidate gene, *golB*, which encodes a copper chaperone protein. The predicted STnc520-*golB* interaction requires further investigation. Additional *in silico* screening further upstream and within the CDS of each of the candidate targets identified some potential non-canonical STnc520 binding sites (Figure 4.12). Furthermore, an additional five genes were predicted to contain STnc520 binding sites by at least two biocomputational target prediction programmes (Table 4.4), and may warrant further investigation, as a result. Studies using mutational analysis and nucleotide exchange of predicted binding sites would be necessary to prove if these sites play a role in STnc520-mediated post-transcriptional gene regulation.

Although no STnc520 targets have been validated as direct targets, it is tempting to speculate about the function of the genes which were differentially expressed upon deletion and over-expression of STnc520, and which serve as promising candidate targets. A 5 minute pulse was sufficient to restore expression of these candidate targets to wild-type levels, and this rapidity may suggest direct regulation of all of these genes by STnc520 (Vogel & Wagner, 2007). However, intracellular levels of sulphur regulate expression of genes within the *cys* regulon (section 4.1.5) and it is possible that

interruption of one step in the pathway may be sufficient to de-regulate genes involved in subsequent steps of the sulphur metabolism pathway (Quan et al., 2002). Expression of the cysB gene which encodes the primary transcriptional activator of the cys regulon is not affected by deletion or over-expression of STnc520, however it cannot be ruled out that STnc520 affects the translation of cysB mRNA, without affecting cysB mRNA stability. A change in the translation efficiency of cysB would, in turn, lead to an altered transcriptional profile of the majority of genes in the CysB regulon, as detected by our RNA-seq-based transcriptomics approach. fdoH and fdnH encode the iron-sulphur subunits of the oxygen and nitrate-inducible formate dehydrogenases. It is, therefore, possible that, like the cys regulon genes, the regulation of fdoH and fdnH expression may also be influenced by the levels of intracellular sulphur. Increased activity of the sulphur reduction pathway, resulting from STnc520-mediated activation of a gene or genes within the cys regulon, may lead to a reduction in available sulphur for the formation of iron-sulphur clusters, while CysB also affects transcription of the fdo operon (Quan et al., 2002). In addition, we speculate that the STnc520-mediated activation of genes involved in the export and sequestration of the toxic metal copper could be a mechanism of protecting the cell against protein damage caused by copper binding or reacting with cellular cysteine (Hodgkinson & Petris, 2012). Overall, the three groups of candidate target genes are intrinsically linked in cellular metabolic processes. We speculate that direct control of expression of one or more of the candidate targets by STnc520 may provide a method of sensing cellular redox levels, via the thiol side-chain of cysteine, and subsequently protecting the cell from the oxidative damage caused by copper, as well as indirectly activating the superoxide dismutase SodCII (Osman et al., 2013), as discussed Repression of the formate dehydrogenase subunits by previously (section 4.1.6). STnc520 could resemble repression of the primary regulator of formate metabolism, fhlA, by the sRNA, OxyS, which is thought to serve as a mechanism of limiting the production of oxygen sensitive formate metabolism systems during an oxidative burst (section 4.1.4) (Altuvia et al., 1997).

Figure 4.13 is a model summarising our current knowledge and hypotheses regarding transcriptional regulation of STnc520, and the putative role of STnc520 within *S*. Typhimurium. At present, our model of STnc520-mediated protection from oxidative stress is purely speculative and further experimentation is required to identify which, if any, of the candidate target genes are directly regulated by STnc520, and by what

mechanism of regulation. Furthermore, experimentation to show that the absence of STnc520 results in increased sensitivity of *S*. Typhimurium to oxidative stress will be necessary to confirm the biological significance of our observations and speculations. If our speculative model for STnc520-mediated gene regulation is correct, however, it indicates that STnc520 could complement the role of other SPI11-encoded genes, in allowing *S*. Typhimurium and other *S*. *enterica* serovars to withstand the stressful intracellular environment during the infection process as previously discussed (section 4.1.3). It is also tempting to speculate that regulation of STnc520 by SprB could, therefore, provide an added layer of cross-talk between the genes involved in cellular invasion and those involved in survival of the intracellular lifestyle.

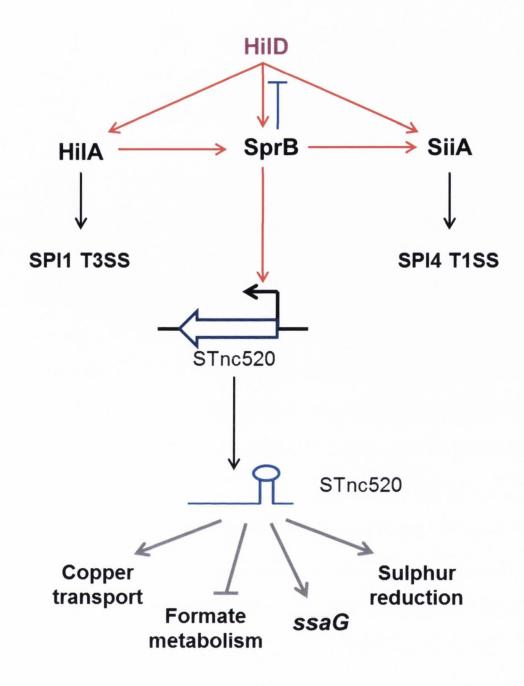


Figure 4.13 Model of the transcriptional regulation and putative role of STnc520

Schematic model summarising our current knowledge and hypotheses regarding the transcriptional regulation of STnc520, and the role of STnc520 within *S.* Typhimurium. STnc520 transcription is directly activated by SprB. SprB also positively regulates expression of the genes encoding the SPI4 Type 1 secretion system, which secretes a large non-fimbrial adhesin, via *siiA* and negatively regulates hilD, resulting in the formation of a negative feedback loop. Absence and over-expression of STnc520 results in the differential expression of genes involved in copper transport, formate metabolism, sulphur reduction and the gene encoding the major subunit of the SPI2 TTSS. The STnc520-mediated regulation of expression of genes involved in these cellular processes occurs via an unknown mechanism. Red arrows denote transcriptional activation. Blue T-bars denote transcriptional repression. Grey arrows and T-bars denote activation and repression, respectively, via an unknown mechanism.

# Chapter 5 Investigating the transcriptional regulation of STnc1480

#### 5.1 Introduction

# 5.1.1 Transcriptional silencing and counter-silencing

Lateral gene transfer in bacteria can be mediated by 3 mechanisms: conjugation, transformation and transduction. Acquisition of new genetic material through lateral gene transfer is the key to microbial evolution and diversity (Arber, 2014). In particular, acquisition of virulence genes or antibiotic resistance determinants has been instrumental in the evolution of bacterial pathogens. For instance, it is estimated that S. Typhimurium LT2 has acquired over 200 DNA regions of greater than 100 bp in length, corresponding to approximately 1400 ORFs, or over one-quarter of its genome, since diverging from its last common ancestor with E. coli approximately 100 million years ago (Porwollik & McClelland, 2003). However there is the potential for foreign DNA to decrease the fitness of the recipient bacteria, either through insertion into a functionally-useful region of the chromosome, or by carrying genes which are toxic to the recipient bacteria. The foreign DNA can also disrupt well-established cellular regulatory networks and gene function (Navarre et al., 2007). Thus, in order to incorporate foreign DNA into the cellular networks, and to allow the stable inheritance of foreign DNA without affecting fitness of the recipient organism, and to provide an evolutionary advantage to the recipient, a mechanism known as "xenogeneic silencing" occurs (Navarre et al., 2007). The nucleoid-associated protein, H-NS, has emerged as a key protein in the mechanism of xenogeneic silencing. H-NS contributes to the overall architecture of the bacterial chromosome and was originally thought of as a general transcriptional silencer that favoured binding to curved regions of DNA (Dorman, 2004), however two ChIP-on-chip based studies determined that the preference for H-NS binding was an AT-rich sequence, such as those found in horizontally-acquired genes. It was concluded that H-NS mainly functions to silence expression of foreign DNA and prevent inappropriate disruption of the expression of ancestral genes (Navarre et al., 2006, Lucchini et al., 2006). H-NS binds DNA and oligomerises, spreading laterally along the DNA (Dorman, 2004). It is thought that H-NS either occludes RNAP from the gene promoter (Lucchini et al., 2006) or that H-NS induces changes in the DNA topology, forming loop structures or bridges, that trap RNAP at the promoter (Stoebel et al., 2008, Navarre et al., 2007), thereby silencing gene expression.

Proteins with full or partial homology to H-NS have been shown to play similar silencing roles to H-NS. StpA is a full-length paralogue of H-NS. The two proteins show the highest level of conservation in the DNA-binding and dimerization domains. This similarity allows StpA and H-NS to form heterodimers. StpA regulates a distinct subset of H-NS-regulated genes (Lucchini et al., 2009). The plasmid-encoded H-NS-like protein Sfh, also regulates a subset of H-NS-regulated genes and acts as a molecular "back-up" for H-NS in an *hns* mutant (Dillon et al., 2010). Hha and its paralogue YdgT are part of a family of proteins that show structural mimicry to the H-NS N-terminal oligomerisation domain and can form complexes with H-NS (Madrid *et al.*, 2007, Ali *et al.*, 2013). Hha silences both SPI1 and SPI2 genes (Fahlen *et al.*, 2001, Silphaduang *et al.*, 2007). Hha and YdgT also play a more general role in silencing laterally transferred DNA (Banos *et al.*, 2009, Aznar *et al.*, 2013, Vivero *et al.*, 2008). It has been suggested that the formation of multimeric complexes between H-NS and chromosomal or plasmid-borne H-NS paralogues, allows H-NS to discriminate between its horizontally-acquired and ancestral gene targets (Banos et al., 2009).

The maintenance of horizontally-acquired DNA in bacterial genomes indicates that this DNA is beneficial to the recipient organism under certain conditions. It follows, therefore, that mechanisms to counteract the silencing effect of H-NS, and H-NS paralogues, must also have evolved to allow expression of the horizontally-acquired genes under appropriate conditions. As previously discussed, counter-silencing is a key method of indirect transcription initiation (Figure 1.2 D). Counter-silencing may be mediated in a protein-independent or protein-dependent fashion. Protein-independent counter-silencing can be brought about by environmental conditions, such as high temperature or osmolarity, that can alter DNA curvature and disrupt H-NS-mediated DNA bridging (Stoebel et al., 2008). In *E. coli*, changes in osmolarity, resulting in changes to DNA superhelicity, have been shown to affect H-NS binding at the *proU* promoter, a locus which encodes proteins involved in the transport of the osmoprotectant glycine betaine (Bouffartigues *et al.*, 2007).

There are many examples of protein-dependent counter-silencing methods. In *Salmonella*, key virulence genes, associated with SPI1 and SPI2, are transcriptionally activated via counter-silencing mechanisms. HilD and HilC activate *hilA* and *rtsA* expression through counter-silencing the repressive effects of H-NS and Hha (Olekhnovich & Kadner, 2006, Olekhnovich & Kadner, 2007). In addition, the

requirement for SsrB for transcription of a number of SPI2 genes was reduced in an *hns* mutant background, suggesting a role for SsrB in counteracting silencing by H-NS at these promoters (Walthers et al., 2007).

SlyA-like proteins have been shown to play an important role in counter-silencing mechanisms. Salmonella SlyA is a winged-helix protein that is related to MarA in E. coli, RovA in Yersinia and PecS in Erwinia. In S. Typhimurium, SlyA has a number of gene targets associated with virulence (particularly SPI2 genes), resistance to antimicrobial peptides and oxidative stress (Navarre et al., 2005, Stapleton et al., 2002, Mutants lacking SlyA are attenuated for intra-macrophage Linehan et al., 2005). survival, adherence to or survival on macrophage cells, resistance to oxidative bursts and resistance to antimicrobial peptides (Libby et al., 1994, Buchmeier et al., 1997, Shi et al., 2004, Stapleton et al., 2002). A subset of genes that are dependent on the PhoP/Q TCS are also dependent on SlyA for their expression. The co-regulated genes are horizontally-acquired PhoP targets, while ancestral PhoP-dependent genes do not require SlyA for transcription (Navarre et al., 2005, Shi et al., 2004). As previously discussed (section 1.2.1), co-regulation by PhoP and SlyA is necessary to antagonise H-NS silencing at the pagC and ugtL promoters (Perez et al., 2008). SlyA does not act as a classical transcriptional activator and the role of SlyA is confined to counter-silencing H-NS, while PhoP is necessary for activation of transcription. H-NS and SlyA were found to have overlapping binding sites in the pagC promoter, however H-NS is not displaced from the pagC promoter by SlyA binding. Rather, both proteins bind the promoter region simultaneously, and SlyA re-models or bends the promoter DNA. The re-modelling of the promoter architecture is likely affect gene transcription by allowing PhoP to bind and recruit RNAP (Perez et al., 2008). There are added layers of feedback control on this system, as both PhoP and SlyA auto-regulate (Stapleton et al., 2002, Soncini et al., 1995) and regulate the expression of each other (Song et al., 2008, Norte et al., 2003, Shi et al., 2004), while H-NS represses expression of phoP (Kong et al., 2008). The layered control of this regulatory mechanism highlights the importance of silencing and subsequent counter-silencing in ensuring the co-ordinated expression of horizontally-acquired genes to maximise recipient bacterial fitness.

In contrast to the role of SlyA as a H-NS antagonist at the *pagC* and *ugtL* promoters in *Salmonella*, an alternative method of counter-silencing by SlyA in *E. coli* has been demonstrated. H-NS and SlyA displace each other, based on the relative abundance of

each protein, at the *hlyE* promoter (Lithgow et al., 2007). This mechanism of mutual antagonism has not yet been demonstrated in *Salmonella*, however it remains possisble that the SlyA protein can act by either mechanism at certain promoters in both species. The SlyA homologue in *Yersinia*, RovA, acts both as a counter-silencer and as a transcriptional activator (Tran *et al.*, 2005). *Salmonella* SlyA could also have a dual function, although transcriptional activation activity has not yet been described for SlyA.

Other mechanisms of factor-dependent antagonism of H-NS mediated silencing have also been reported. For instance, the LeuO protein acts as a "boundary element" to prevent the spread of H-NS nucleoprotein filaments at the *leuO* promoter (Chen *et al.*, 2005, Chen & Wu, 2005). In addition to counter-silencing at the *leuO* promoter, a ChIP-on-chip based study found a large overlap between genes simultaneously bound by both LeuO and H-NS, suggesting LeuO is an important protein in antagonising H-NS-mediated gene silencing on a global scale (Dillon *et al.*, 2012). In *E. coli*, translation of *hns* mRNA is negatively regulated by the sRNA DsrA (Lease *et al.*, 1998) and DsrA has, thus been shown to act as a counter-silencer of H-NS at the *rcsA* promoter (Sledjeski & Gottesman, 1995). The DsrA-mediated counter-silencing effect is likely to be due to the reduction in levels of H-NS protein within the cell. Finally, other nucleoid associated proteins such as HU and Fis, as well as H-NS itself, can antagonise H-NS-mediated gene silencing (Stoebel et al., 2008).

The abundance of chromosomal and plasmid-borne H-NS and H-NS-like proteins in enteric bacteria highlights the importance of NAPs in regulating expression of ancestral chromosomal genes as well as genes acquired horizontally on mobile genetic elements. The fact that the stable inheritance of laterally-acquired DNA occurs relatively rarely and at a low frequency (Navarre et al., 2007), demonstrates the difficulties involved in integrating this DNA into existing cellular networks without compromising the fitness of the recipient organism, while providing an evolutionary benefit for the recipient. The gene silencing methods developed by recipient bacteria share the common theme of usually involving nucleoid-associated proteins, such as H-NS or H-NS paralogues, while a variety of mechanisms and factors involved in transcriptional counter-silencing have apparently evolved. It has been suggested that these counter-silencing mechanisms have evolved in this bespoke way to meet the regulatory needs of the bacterium under particular conditions, allowing for beneficial foreign DNA to become rapidly co-opted into existing regulatory networks as necessary (Stoebel et al., 2008).

#### 5.1.2 Trans-acting sRNAs that activate gene expression

As shown in Figure 1.4, sRNAs display a variety of different mechanisms of post-transcriptional gene regulation, which typically involve the sRNA acting to decrease the efficiency of translation or decrease the stability of an mRNA transcript (Vogel, 2009). However, examples of sRNAs acting as positive regulators of gene expression have also been reported (Frohlich & Vogel, 2009).

sRNA-mediated positive gene regulation typically involves an "anti-antisense" mechanism. Inhibitory secondary structures surrounding the RBS, in the extended RBS, or within the first few codons of an mRNA transcript, can lead to a reduction in the efficiency of translation initiation and result in a decrease in protein synthesis. These sequestering secondary structures are common in bacterial mRNAs and often underlie mechanisms of cis-acting ribo-regulation. Base-pairing of the sRNA can eliminate the repressive effects of the secondary structure (Frohlich & Vogel, 2009). The best-studied example of sRNAs playing a positive role in gene expression is the activation of RpoS expression by DsrA, RprA and ArcZ in E. coli (Lease et al., 1998, Majdalani et al., 2001, Mandin & Gottesman, 2010). The long 5' UTR of the rpoS mRNA forms a 100 nt inhibitory hairpin structure that sequesters the RBS and prevents efficient translation of rpoS under non-stress conditions. As previously discussed (section 1.3.2), the sRNAs, DsrA, RprA and ArcZ, are each induced under different stress conditions, and bind the stem of the hairpin structure to open it and alleviate the repression of rpoS translation (Majdalani et al., 1998, Majdalani et al., 2002, Mandin & Gottesman, 2010). Binding of DsrA and RprA to the rpoS 5' UTR sequence increases stability of the rpoS mRNA by protecting the mRNA from RNase E-mediated degradation (McCullen et al., 2010).

Another example of sRNA-mediated positive control of gene expression is the regulation of *shiA* mRNA by the Fur-regulated RyhB sRNA in *E. coli*, which also occurs by an anti-antisense mechanism. Binding of RyhB to the 5' UTR of *shiA* prevents the formation of an intrinsic inhibitory structure in this region and increases the stability of the *shiA* transcript, leading to an increase in ShiA protein synthesis (Prevost *et al.*, 2007).

The homologous sRNAs GlmY and GlmZ post-transcriptionally activate expression of *glmS*, the gene encoding glucosamine-6-phosphate synthase in *E. coli*. However, only GlmZ directly activates *glmS* expression, via an anti-antisense mechanism, by freeing the *glmS* RBS for ribosome access and efficient translation. On the other hand, GlmY

indirectly activates *glmS* expression by overcoming the repressive effect of the YhbJ protein, which inactivates GlmZ through 3' end processing and loss of the *glmS* binding site (Urban & Vogel, 2008).

Recently, a translation-independent method of gene activation, through unmasking of an RNase E binding site, was demonstrated for the sRNA, RydC (Frohlich et al., 2013). A number of methods of sRNA-mediated indirect activation of gene expression, such as mRNA traps, and *cis*-encoded positive ribo-regulators exist (Frohlich & Vogel, 2009). Interestingly the sRNAs discussed here are not limited to acting as positive regulators of gene expression and have been shown to act as negative regulators also. For example, DsrA inhibits *hns* translation and RyhB represses translation of a number of gene targets (Lease et al., 1998, Masse & Gottesman, 2002, Masse et al., 2005). This dual ability highlights the adaptability of sRNAs in post-transcriptional gene regulatory networks because the same type of base-pairing interaction, requiring Hfq, may be used for activation or repression of mRNA targets.

## 5.2 Results

#### 5.2.1 Characterisation of STnc1480

STnc1480 was first identified following RNA-seq analysis of wild-type 4/74 grown to ESP (Kröger et al., 2012). STnc1480 is a relatively long sRNA at approximately 395 nt. Most sRNAs are between 50 and 250 nt in length (Vogel, 2009), however other long sRNAs have been reported, such as the 1200 nt AmgR (Lee & Groisman, 2010) and IsrH\_1\_2 and IsrA, which are approximately 450 and 420 nt in length respectively (Padalon-Brauch et al., 2008). STnc1480 is encoded in an intergenic region on the opposite strand to *yeaH* and *yeaJ*, as shown in Figure 5.1 A. STnc1480 was not considered to be Hfq-bound in the recent classification (Kröger et al., 2013), based on Hfq co-immunoprecipitation under 7 environmental conditions (Chao et al., 2012). Kröger *et al* used a stringent cut-off to identify the sRNAs which are most highly associated with Hfq (enrichment factor >5 for Hfq co-immunoprecipitation, compared to a mock co-immunoprecipitation). There is approximately 3-fold enrichment for STnc1480 binding to Hfq, indicating that STnc1480 does associate with Hfq but the interaction may be weak under the conditions tested.

To further characterise STnc1480, a stability experiment was carried out to investigate the stability of the STnc1480 transcript in wild-type and  $\Delta hfq$  cells. In addition to the role of Hfq in facilitating and stabilising sRNA interactions with mRNA targets, Hfq also binds and protects sRNAs from degradation by cellular ribonucleases prior to target binding, thus many sRNAs require Hfq for stability (Vogel & Luisi, 2011). STnc1480 was recently shown to be most highly expressed under conditions that mimic the intracellular macrophage environment (Kröger et al., 2013) and within murine macrophages (Srikumar et al., 2014), so wild-type and  $\Delta hfq$  strains were grown in SPI2-inducing PCN media to an OD<sub>600</sub>0.3, cells were harvested for RNA extraction from both strains (Time 0) and Rifampicin was added to a final concentration of 150 µg/mL to stop cellular transcription, as previously described (section 2.6.10). Cells were harvested for RNA extraction at the time points, indicated in Figure 5.1 B, following the addition of Rifampicin, and total RNA was probed for the abundance of the STnc1480 transcript by northern blot (as described in section 2.6.6). In wild-type cells STnc1480 has a half-life of approximately 3 minutes. STnc1480 fails to accumulate in a  $\Delta hfq$  mutant (Figure 5.1 B). Densitometry calculations were done using ImageJ software based on northern blots of wild-type RNA from 2 independent biological experiments (Figure 5.1 C). It was not possible to estimate the effect of Hfq on STnc1480 stability using this experiment. It is likely that STnc1480 requires Hfq for protection from cellular ribonucleases but it cannot be ruled out that the absence of the Hfq proteins affects the steady state levels of STnc1480. Expression of phoP is reduced approximately 8-fold in the absence of Hfq and as previously discussed STnc1480 expression is PhoP-dependent (Figure 3.6 E). Therefore, the reduction in STnc1480 expression in the  $\triangle hfq$  mutant may be due to reduced PhoP expression.

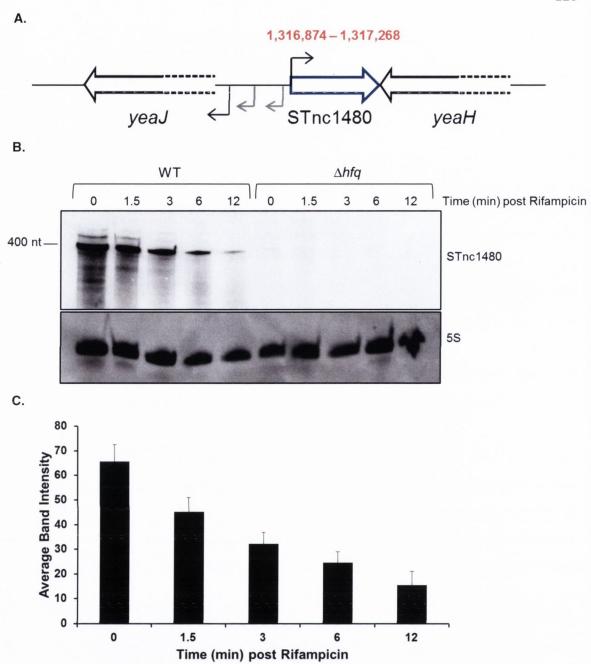


Figure 5.1 Chromosomal location and stability of STnc1480

**A.** Schematic showing the chromosomal location of the STnc1480 gene. STnc1480 (white arrow with blue outline) is encoded on the positive strand and is located in the intergenic region between yeaJ and yeaH (white arrows with black outline). Bent arrows denote TSS, black arrows are primary TSS and smaller grey arrows are secondary TSS (Kröger et al., 2013). Arrows show the direction of transcription of the indicated genes. Dotted lines indicate the genes are not drawn to scale. The chromosomal co-ordinates of STnc1480 are written in red. **B.** Representative northern blot from two independent biological experiments showing the abundance of the STnc1480 transcript in wild-type and  $\Delta hfq$  cells, following the inhibition of transcription by addition of Rifampicin (section 2.6.10). Numbers correspond to the time (in minutes), following addition of Rifampicin, when cells were withdrawn for RNA extraction. 5S RNA was probed as a loading control. **C.** Densitometry measurement based on the intensity of each northern blot band, corresponding to wild-type STnc1480 from 2 independent experiments, measured using ImageJ software. Error bars are based on the standard deviation of the mean from 2 independent experiments.

# 5.2.2 Predicted structure and conservation of STnc1480

As shown in Figure 3.12, STnc1480 is a Salmonella-specific sRNA. Typhimurium STnc1480 gene shares between 80-90% sequence identity with regions in S. Typhi and S. Paratyphi B. However, a sequence sharing only 25% sequence identity with S. Typhimurium STnc1480 is present in S. bongori (Figure 3.12). Regions of high sequence identity within sRNAs are often functionally important for the sRNA. For example, seed regions for target binding must be highly conserved to maintain regulatory function of the sRNA (Altuvia et al., 1998). A multiple alignment of the STnc1480 sequence from 4/74 with other S. enterica serovars and with the less well-conserved sequence from S. bongori was carried out, using Multalin software (section 2.10.3), in order to identify regions of functional importance of STnc1480, based on sequence identity (Figure 5.2 A). The full STnc1480 gene sequence is present in other S. Typhimurium strains, such as the invasive non-typhoidal African strain D23580, and also in the Paratyphi B serovar. The entire gene sequence is present in S. Typhi but the S. Typhi sequence also contains a 9 nt insertion from nucleotide 148 to 157. The STnc1480 sequence is only partially conserved in S. Enteritidis and S. bongori. The first 128 nucleotides are the only part of the sRNA that is present in S. Enteritidis, while the sequence is not well conserved after nucleotide 257 in S. bongori. Therefore we speculate that the first 128 nt of the gene are likely to contain the functionally important sequences of STnc1480 as they are well-conserved within the Salmonella genus.

STnc1480 is predicted to contain a small ORF from nucleotide 52 to 249. There is a close-to-consensus ribosome binding site (CGGAGG) upstream of the putative start codon of the small ORF. The putative start codon is conserved in all of the species and serovars presented in Figure 5.2 A. However, in *S.* Enteritidis the putative RBS is more degenerate from the consensus sequence (CTGAGG), while the RBS is consensus (AGGAGG) in *S. bongori*. The predicted coding sequence within STnc1480 should produce a small protein of approximately 7.3 kDa, however in our hands, 4/74 STnc1480 does not express a small protein under the conditions tested, following tagging of the putative ORF and western blot analysis, as well as global proteomic analysis (data not shown). STnc1480 was, therefore, considered a non-coding RNA for this study, and the putative protein-coding functions of STnc1480 were not investigated further.

Software available from <a href="www.nupack.org">www.nupack.org</a> was used to predict the secondary structure of STnc1480 (Zadeh et al., 2011). As shown in Figure 5.2 B, STnc1480 is predicted to have a complex structure, consisting of a number of stem structures with both internal and terminal bulge and loop structures. The colours of the bases relate to the probability that they base-pair as predicted. The most stable and highly probable structure is the predicted p-independent terminator extending from nucleotide 362 to 395. Other loops and bulges could potentially be involved in target binding due to the lack of base-pairing within the structures, and thus the ease at which the structure could be opened. Loop structures within the conserved first 128 nt are of particular interest as regions of functional importance for STnc1480.

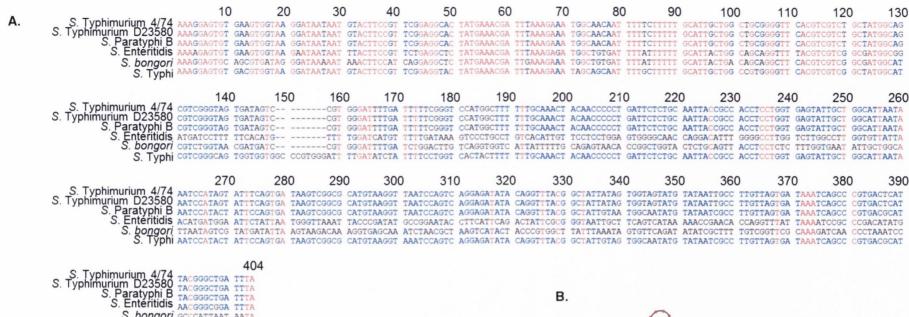
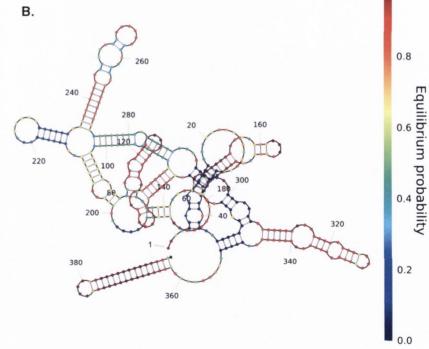


Figure 5.2 Conservation and predicted structure of STnc1480

S. bongori geceattaat aata S. Typhi taegggetga tita

**A.** Multiple alignment of the STnc1480 sequence from the indicated strains in the *Salmonella* genus (*S.* Typhimurium 4/74, *S.* Typhimurium D23580, *S.* Paratyphi B, *S.* Enteritidis, *S. bongori*, *S.* Typhi). Numbers correspond to the position of that nucleotide relative to the TSS (+1) of the STnc1480 gene. Red indicates >90% sequence identity, Blue indicates <90% >50% sequence identity, black indicates <50% sequence identity across the six genomes. **B.** Software from www.nupack.org was used to predict the structure of the STnc1480 sRNA. The colours of each base-pair indicate the probability that this base-pairing interaction occurs. The heatmap to the right of the predicted structure indicates the colour used for each probability score. STnc1480 is predicted to form a complex structure with a minimum free energy of -99.8 kcal/mol. Numbered nucleotides indicate the position of that nucleotide relative to the TSS (+1) of the STnc1480 gene, and can be directly compared to the numbers in the sequence alignment shown in panel (**A**).



Free energy of secondary structure: -99.80 kcal/mol

# 5.2.3 STnc1480 has a SPI2-like pattern of expression

As previously discussed, regulation of STnc1480 resembles SPI2-associated genes and STnc1480 demonstrates a high level of correlation with the archetypical SPI2 gene ssaG based on its pattern of expression in a panel of 15 regulatory mutants and wild-type grown under 5 environmental conditions (Pearson correlation coefficient: 0.91) (section 3.2.10). Comparison of STnc1480 expression to ssaG expression from RNA-seq data of wild-type 4/74 grown under 22 environmental conditions (Kröger et al., 2013) and within murine macrophages (Srikumar et al., 2014) also shows that STnc1480 has a SPI2-like pattern of expression in the wild-type strain (Pearson correlation coefficient: 0.88). As shown in Figure 5.3 A, STnc1480 is not highly expressed during growth in rich media, although STnc1480 expression does become slightly up-regulated at ESP, resembling other SPI2-associated genes, as a result of HilD-mediated cross-talk between SPI1 and SPI2 (Bustamante et al., 2008). However, STnc1480 expression and SPI2 gene expression is highest under conditions which mimic the intracellular environment. Expression of STnc1480 is significantly higher in low magnesium conditions. Reduction of the levels of magnesium from 1 mM to 10 µM in SPI2-inducing PCN results in an approximately 5-fold increase in STnc1480 expression. The magnesium ion-dependent increase in STnc1480 levels is one of the main differences in expression levels between STnc1480 and ssaG expression, as ssaG is not significantly affected by changes in magnesium levels. STnc1480 is not as highly induced as ssaG following peroxide and nitric oxide shock. STnc1480 and ssaG are both highly expressed within murine macrophages 8 hours post-infection (Figure 5.3 A).

As previously discussed, STnc1480 has multiple regulatory inputs (Figure 3.6 E and Figure 5.3 B and C). Expression of STnc1480 decreases in the SPI1-associated mutants, as does ssaG. Again, this may reflect the HilD-mediated cross-talk between SPI1 and SPI2 and could suggest that there are further factors involved in coordinate regulation of both of the main pathogenicity islands. Unlike ssaG, however, STnc1480 expression decreases in the absence of  $\sigma^{38}$  at LSP. Expression of ssaG is up-regulated in the absence of  $\sigma^{38}$ . However the decrease in STnc1480 expression in the absence of  $\sigma^{38}$  appears to be indirect as the STnc1480 promoter does not contain many of the hallmarks of a  $\sigma^{3}$ -dependent promoter (Typas et al., 2007), in fact the STnc1480 -10 and -35 hexamers are close to consensus for the previously defined  $\sigma^{70}$ -dependent recognition and

 $\sigma^{70}$ -mediated transcription initiation (Kröger et al., 2012). Expression of STnc1480 is up-regulated in the absence of FNR, an effect that was previously observed for SPI2-associated genes (section 3.2.3 I).

STnc1480 expression is reduced approximately 6 to 7-fold in many of the SPI2-associated regulatory mutants, and reduced approximately 60-fold in the  $\Delta slyA$  and  $\Delta phoP$  mutants. A signal corresponding to STnc1480 is not detectable in either of these mutant strains by northern blot (Figure 3.6 E), indicating that STnc1480 requires both of these proteins for its expression. As shown in Figure 3.11 D, STnc1480 is grouped with SPI2-associated genes according to the pattern of regulation of STnc1480. In particular STnc1480 expression resembles the subset of genes, such as ugtL and pag genes, that require both PhoP and SlyA to counter-silence H-NS and activate their transcription (Perez et al., 2008). The pattern of regulation of STnc1480, coupled with the up-regulation of STnc1480 under low  $Mg^{2+}$  conditions, which is one of the key signals known to activate the PhoP regulon (Garcia Vescovi et al., 1996), led us to investigate if activation of STnc1480 expression requires PhoP and SlyA to antagonise silencing H-NS.

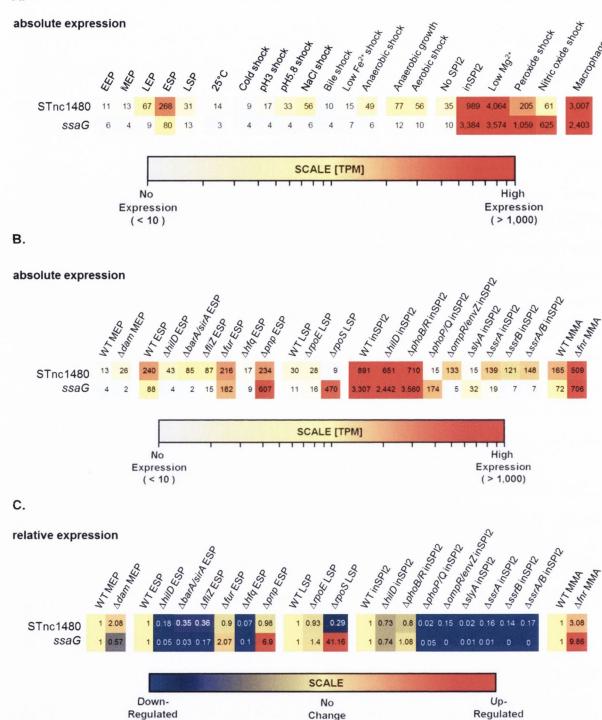


Figure 5.3 STnc1480 has a SPI2-like pattern of expression and regulation

>6-fold

**A.** Screenshot from the SalCom website showing absolute expression (TPM values) of STnc1480 and the archetypical SPI2 gene ssaG under 20 environmental conditions and within murine macrophages (Kröger et al., 2013, Srikumar et al., 2014). Expression of both genes is induced under conditions that mimic the intracellular environment and within the intracellular environment. **B.** Absolute expression and **C.** Relative expression of the STnc1480 gene in 18 regulatory mutants and their wild-type comparators (this study). STnc1480 is differentially expressed (>2.5-fold) in  $\Delta hilD$ ,  $\Delta barA/sirA$ ,  $\Delta fliZ$ ,  $\Delta hfq$ ,  $\Delta rpoS$ ,  $\Delta phoP/Q$ ,  $\Delta ompR/envZ$ ,  $\Delta slyA$ ,  $\Delta ssrA/B$  and  $\Delta fnr$  mutants.

>6-fold

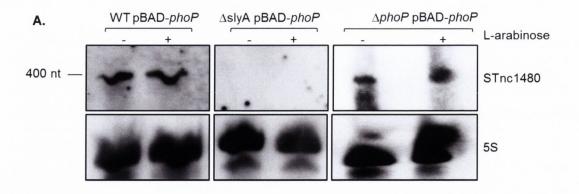
# 5.2.4 STnc1480 requires PhoP and SlyA for expression

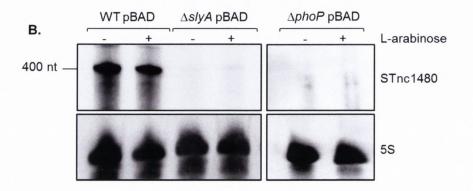
To investigate the role of PhoP and SlyA in regulating expression of STnc1480, the genes encoding both proteins were cloned into the multiple cloning site (MCS) of the pBAD vector, placing both genes under the control of the arabinose-inducible P<sub>BAD</sub> promoter. The pBAD-phoP and pBAD-slyA strains were used to express PhoP and SlyA ectopically in wild-type and various mutant backgrounds. As previously described (section 2.2.3), all strains were grown to an OD<sub>600</sub>0.2 under inSPI2 conditions, the cultures were split and L-arabinose was added to one half, while the other half was left un-induced. Cultures were grown for a further 45 minutes before transcription was stopped and cells were harvested for RNA extraction, as previously described (Pfeiffer et al., 2007). Glucose was already present in the culture medium at a final concentration of 0.4% so further glucose was not added despite a level of leakiness from the P<sub>BAD</sub> promoter in the un-induced samples (Figure 5.4 A and C). To ensure the specificity of the results, control strains carrying an empty pBAD vector were treated in the same way (Figure 5.4 B and C). Total RNA was probed for STnc1480 expression by northern blot.

Ectopic expression of PhoP in a  $\Delta slyA$  background does not result in restoration of STnc1480, while ectopic expression of PhoP in  $\Delta phoP$  background does result in a slight but specific restoration of STnc1480 expression (Figure 5.4 A). The effect is specific as addition of L-arabinose does not restore STnc1480 expression in the  $\Delta phoP$  mutant carrying the empty pBAD vector (Figure 5.4 B).

Ectopic expression of SlyA in a wild-type background causes further accumulation of STnc1480 than in a wild-type strain carrying an empty pBAD vector. Exogenously expressed SlyA specifically rescues STnc1480 expression in a  $\Delta slyA$  background. Addition of L-arabinose does not restore STnc1480 expression in the  $\Delta slyA$  mutant carrying an empty pBAD vector. However, as is the case for exogenous expression of PhoP in a  $\Delta slyA$  background, ectopically expressed SlyA is unable to rescue STnc1480 expression in a  $\Delta phoP$  background.

Expression of PhoP and SlyA for 45 minutes is sufficient to rescue expression of STnc1480 in their isogenic mutant backgrounds; however, neither protein can restore expression of STnc1480 in the absence of the other protein. This shows that both proteins are necessary for optimal expression of STnc1480, strengthening the hypothesis that PhoP and SlyA act together to counteract a H-NS-mediated silencing effect.





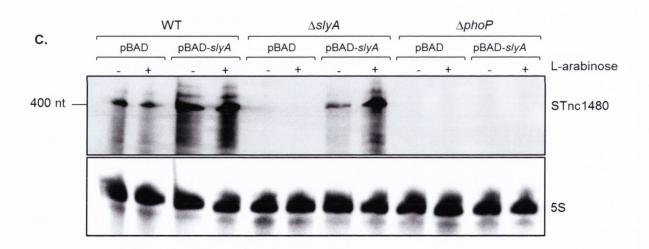


Figure 5.4 Analysis of STnc1480 expression following ectopic expression of PhoP and SlyA

**A.** Northern blot showing that STnc1480 is expressed in wild-type cells but not in a  $\Delta slyA$  mutant when PhoP expression is induced from the P<sub>BAD</sub> promoter. STnc1480 expression is restored when PhoP is exogenously expressed in a  $\Delta phoP$  background. **B.** The effect of exogenously expressed PhoP is specific as there is no restoration of STnc1480 expression following addition of L-arabinose in a  $\Delta phoP$  mutant carrying an empty pBAD vector. **C.** Northern blot showing that STnc1480 is expressed in wild-type cells and accumulates to a higher level when SlyA is ectopically expressed from the P<sub>BAD</sub> promoter. STnc1480 expression is specifically restored when SlyA is ectopically expressed in a  $\Delta slyA$  mutant, and addition of L-arabinose to  $\Delta slyA$  carrying an empty pBAD vector does not result in STnc1480 restoration. STnc1480 expression is not restored by ectopic induction of SlyA expression in a  $\Delta phoP$  mutant background. -/+ indicates the respective absence or presence of the inducer L-arabinose. All northern blots were probed for 5S as a loading control.

# 5.2.5 Silencing of STnc1480 by H-NS under non-inducing conditions

The nucleoid associated protein H-NS has a preference for binding DNA with the higher AT content of horizontally-acquired DNA (Lucchini et al., 2006, Navarre et al., 2006). The average AT content of the ancestral *S.* Typhimurium chromosome is approximately 47.8%, while the AT content of STnc1480 and its promoter region is 61.7%. The high AT content and the lack of conservation of the STnc1480 gene sequence outside of the *Salmonella* genus are hallmarks of horizontal gene transfer, and suggest that STnc1480 could be an ideal target gene for H-NS mediated silencing.

As shown in Figure 5.3 A, STnc1480 is not highly expressed during exponential growth in rich medium. Wild-type and the *hns*-1::*kan* strain (JH3774), which contains a mutation in the DNA binding domain of H-NS (Falconi et al., 1991, Hinton et al., 1992), were grown to MEP in Lennox medium and RNA was extracted. Total RNA was probed for STnc1480 expression by northern blot. STnc1480 expression is de-repressed under non-inducing conditions in the absence of a functional H-NS protein (Figure 5.5 A).

To investigate if H-NS binds the STnc1480 promoter and directly silences STnc1480 expression under non-inducing conditions, chromatin immunoprecipitation coupled with quantitative real-time PCR (ChIP-qPCR) was performed using hns-3xFLAG::kan (JH3777). The ChIP assay was carried out as previously described (section 2.8) and qPCR was carried out as previously described (section 2.6.4). The proV promoter and the hemX gene were used as positive and negative controls respectively as H-NS binds to the proV promoter and does not bind to the hemX gene (Dillon et al., 2010). The amplified ChIP DNA was normalised to amplified Input DNA, which was extracted prior to immunoprecipitation (IP/Input). Figure 5.5 B is representative of two independent biological replicates and shows that, H-NS associates with the STnc1480 promoter, with significantly greater enrichment observed for the experimental IP sample (FLAG), compared to the mock IP sample. Furthermore, the positive control, proV, was enriched for H-NS binding, while the negative control, hemX, displayed little enrichment in the experimental ChIP DNA sample, compared to the mock ChIP sample. Figure 5.5 C shows greater enrichment of the STnc1480 (and proV) promoters compared to the negative control region, following subtraction of the background levels of binding, as quantified by the mock IgG control. These data indicate that H-NS binds and directly silences expression from the STnc1480 promoter under non-inducing conditions.

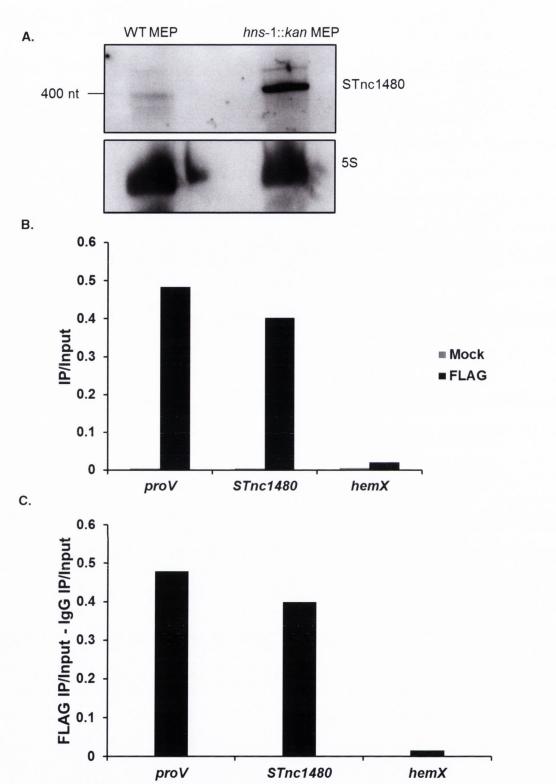


Figure 5.5 H-NS binds the STnc1480 promoter and silences STnc1480 transcription

**A.** Northern blot showing STnc1480 is not highly expressed in wild-type cells grown to MEP in rich medium but STnc1480 expression is de-repressed in a *hns* mutant grown under the same conditions. 5S RNA was probed as loading control. **B.** qPCR data of representative ChIP assay from 2 independent experiments demonstrating direct binding of H-NS to the STnc1480 promoter under non-inducing conditions. Experimental (FLAG) and mock ChIP DNA from the STnc1480 promoter and the positive and negative control regions, *proV* and *hemX* respectively, was normalised to the starting amount of DNA (IP/Input). **C.** qPCR data of representative ChIP assay from 2 independent experiments shows enrichment for H-NS binding of the STnc1480 and positive control *proV* promoters, compared to the negative control region *hemX* following subtraction of background mock ChIP DNA (as described in section 2.8.5).

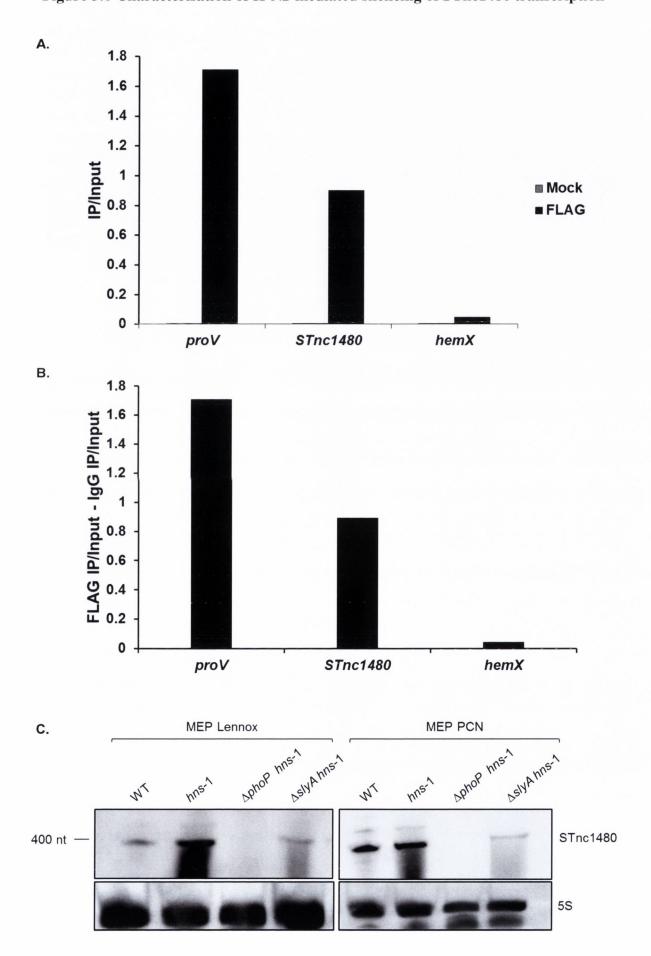
# 5.2.6 Characterisation of the H-NS-mediated silencing of STnc1480 transcription

To investigate if H-NS remains bound at the STnc1480 promoter under conditions when STnc1480 is highly expressed, the ChIP assay was repeated under Mg<sup>2+</sup>-limiting conditions and ChIP DNA was analysed by qPCR, as before. The *hns*-3xFLAG strain was grown to OD<sub>600</sub>0.3 in low Mg<sup>2+</sup> PCN and ChIP was carried out, as previously described (section 2.8). The *proV* promoter and *hemX* gene were used as positive and negative control regions. Figure 5.6 A is representative of two independent biological replicates and shows that the STnc1480 promoter demonstrates greater enrichment for H-NS binding in the experimental (FLAG) ChIP DNA than the mock control DNA. The *proV* promoter is also enriched for H-NS binding, while *hemX* DNA was not enriched in the experimental ChIP sample compared to the mock ChIP sample. The *proV* and STnc1480 promoters are still highly enriched for H-NS binding, compared to *hemX*, following subtraction of the mock ChIP DNA (Figure 5.6 B). These data indicate that H-NS remains bound at the STnc1480 promoter under conditions where STnc1480 expression is highly induced, suggesting that PhoP and/or SlyA do not displace H-NS from the STnc1480 promoter to induce transcription of STnc1480.

To further understand the relationship between STnc1480 and the TFs PhoP and SlyA, and to characterise the roles of PhoP and SlyA in the counter-silencing mechanism, we investigated the expression of STnc1480 in strains lacking a functional H-NS protein in combination with either  $\Delta phoP$  or  $\Delta slyA$  mutants. These strains were grown under non-inducing and inducing conditions for STnc1480 and total RNA was probed for STnc1480 expression by northern blot. In the absence of silencing by H-NS, under both non-inducing and inducing conditions, STnc1480 no longer requires SlyA for transcription. However STnc1480 still requires PhoP for transcription under both conditions (Figure 5.6 C). Expression from the STnc1480 promoter in the hns-1::kan  $\Delta slyA$  background is not as high as in wild-type or the single hns-1::kan mutant. I speculate that the lower level of STnc1480 expression may be due to a reduction in phoP expression in the  $\triangle slyA$  mutant (>4-fold; Appendix II). Positive regulation of phoP expression by SlyA has been previously reported (Song et al., 2008). These results indicate that SlyA is only necessary to activate transcription from the STnc1480 promoter in the presence of H-NS, suggesting that the main role of SlyA is to counter-silence the repressive effect of H-NS. As H-NS remains bound at the STnc1480 promoter under conditions that induce STnc1480 expression, it can be speculated that SlyA mediates

counter-silencing via a promoter re-structuring mechanism that makes the promoter more accessible for RNAP recognition, RNAP binding or transcription initiation, as has been reported for the pagC promoter (Perez et al., 2008) The fact that STnc1480 is not expressed in a  $\Delta phoP$  mutant, even in the absence of H-NS, indicates that PhoP is required for activation of STnc1480 transcription, perhaps through recruitment of RNAP, or through aiding of promoter recognition and binding.

Figure 5.6 Characterisation of H-NS mediated silencing of STnc1480 transcription



#### Figure 5.6 Characterisation of H-NS-mediated silencing of STnc1480 transcription

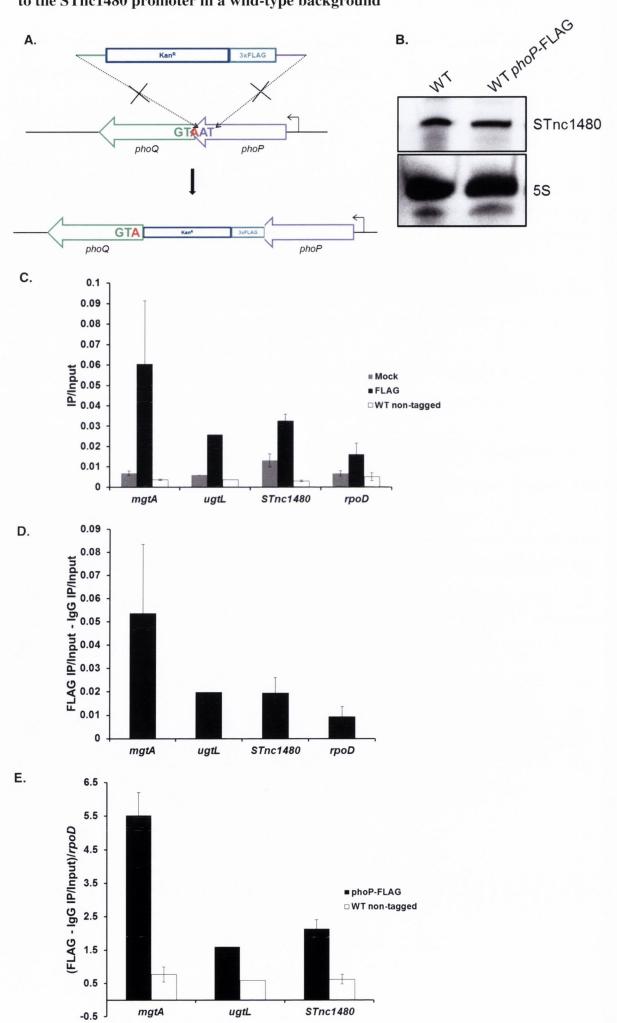
**A.** qPCR data of representative ChIP assay from 2 independent experiments demonstrates direct binding of H-NS to the STnc1480 promoter under inducing, low Mg<sup>2+</sup> conditions. Experimental (FLAG) and mock immunoprecipitated DNA from the STnc1480 promoter and the positive and negative control regions *proV* and *hemX*, respectively, was normalised to the starting amount of DNA, which was extracted prior to immunoprecipitation (IP/Input). **B.** qPCR data of representative ChIP assay from 2 independent experiments shows enrichment for H-NS binding of the STnc1480 and *proV* promoters, compared to the negative control region *hemX*, following subtraction of background mock immunoprecipitated material (as described in section 2.8.5). **C.** Northern blot showing expression of STnc1480 under non-inducing and inducing conditions in wild-type and *hns* mutant backgrounds combined with *slyA* or *phoP* mutations. 5S RNA was probed as a loading control.

# 5.2.7 Binding of PhoP to the STnc1480 promoter

To confirm that PhoP is necessary for direct activation of STnc1480 expression, 3 FLAG tags were added to the *phoP* gene to generate a C-terminally tagged PhoP protein (JH3775) (as described in section 2.5.2). Due to the fact that the *phoQ* gene, which encodes the sensor kinase of the PhoP/Q TCS, shares the *phoP* TSS and that the start codon of *phoQ* overlaps the stop codon of *phoP*, it was important not to disrupt expression of *phoQ* by addition of the tag to the *phoP* gene. The native start codon for *phoQ* was re-instated by the forward oligonucleotide of the oligonucleotide pair that was used to tag the *phoP* gene (Appendix IX) to avoid disrupting translation initiation of *phoQ*. Insertion of the DNA fragment, containing the FLAG epitope tag and selective marker, deleted 2 nt of the *phoP* stop codon and inserted the tag in-frame with the *phoP* start codon, while maintaining the start codon for *phoQ* (Figure 5.7 A). The functionality of the tagged PhoP strain was assessed based on the expression of STnc1480. Wild-type and *phoP*-3xFLAG were grown in low Mg<sup>2+</sup> PCN, total RNA was extracted and probed for expression of STnc1480 by northern blot. STnc1480 expression is not significantly affected by addition of the FLAG tag to the *phoP* gene (Figure 5.7 B).

Preliminary experiments demonstrated that the level of PhoP enrichment, even at published PhoP-bound promoters, was low. For this reason a wild-type (non-tagged) strain was analysed by ChIP with the anti-FLAG antibody in parallel with *phoP-3xFLAG*, as a negative control, to show the specificity of the ChIP reaction although the enrichment levels are low. The tagged-*phoP* and non-tagged-*phoP* strains

were grown in low Mg<sup>2+</sup> PCN to an OD<sub>600</sub>0.3. Cells were harvested from both strains and the ChIP assay was performed, as previously described (section 2.8), on two independent occasions. A mock control was not included in the immunoprecipitation reaction for the wild-type control strain. The mgtA (2 biological replicates) and ugtL (1 biological replicate) promoters were used as positive control promoters as PhoP binding in these regions has previously been demonstrated, while the rpoD promoter (2 biological replicates) was used a negative control region as PhoP does not bind the rpoD promoter (Shin & Groisman, 2005, Perez et al., 2008). qPCR was used to analyse the ChIP DNA. Figure 5.7 C shows the experimental (FLAG and WT non-tagged) and mock ChIP DNA, normalised to Input DNA, from each strain, at all promoters tested (IP/Input). There is a greater enrichment of the STnc1480 promoter observed for the FLAG ChIP compared to the mock ChIP, and compared to the non-tagged control strain, than at the negative control region. Figure 5.7 D shows the level of PhoP enrichment at each promoter when the mock ChIP DNA is subtracted. The STnc1480 promoter DNA is approximately 2-fold enriched compared to the negative control region, following subtraction of the mock ChIP DNA. It is evident from these data that, like the PhoP-bound mgtA and ugtL promoters, the STnc1480 promoter, is enriched for PhoP binding. Due to the biological variability from two independent biological replicates, the negative control rpoD promoter was used as an internal qPCR control for the experimental ChIP DNA from both the phoP-3xFLAG and non-tagged strains (with background subtracted) (Figure 5.7 These data indicate that PhoP and specifically associates with the STnc1480 promoter.



# Figure 5.7 Epitope tagging of PhoP and investigation of PhoP binding to the STnc1480 promoter in a wild-type background

A. Schematic showing the construction of JH3775 by the addition of 3 FLAG tags to the 3' end of the *phoP* gene, as previously described (section 2.5.2). The *phoP* stop codon overlaps the phoQ start codon by one nt (A; indicated in red). Integration of the epitope tag (teal) and kanamycin resistance cassette (blue) by λ-Red-mediated recombination of oligonucleotide sequences homologous to the regions of DNA flanking the tagging site, resulted in the removal of 2 nt from the phoP stop codon (TA; indicated in purple), while re-instating the first nt of the phoQ start codon. Homologous recombination is denoted by Xs. Coloured regions flanking the FLAG tag and Kan<sup>R</sup> cassette denote DNA regions homologous to the regions flanking the tag insertion site. B. Northern blot showing STnc1480 expression is not affected in a strain expressing FLAG-tagged PhoP. C. qPCR data of ChIP assay from two independent experiments demonstrates direct binding of PhoP to the STnc1480 promoter under inducing low Mg<sup>2+</sup> conditions. Experimental and ChIP DNA from the STnc1480 promoter and the positive control promoters (mgtA and ugtL) and the negative control promoter (rpoD) from phoP-3xFLAG and a non-tagged control strain were normalised to the starting amount of DNA (IP/Input). D. qPCR data of ChIP assay from two independent experiments showing enrichment for PhoP binding of the STnc1480 and positive control promoters compared to the negative control region following subtraction of the mock ChIP DNA. E. ChIP DNA from the phoP-3xFLAG and wild-type non-tagged strains was internally normalised to the negative rpoD promoter. Error bars are based on the standard deviation from 2 independent experiments.

#### 5.2.8 ChIP analysis of PhoP binding in a hns mutant background

Enrichment of the *mgtA* promoter for PhoP binding is much higher than at the other published PhoP binding sites and at the STnc1480 promoter. The *phoP*, *pagC*, *ugtL* and STnc1480 promoters are bound and their transcription is silenced by H-NS, while *mgtA* is an ancestral gene which is not silenced by H-NS (Perez et al., 2008, Kong et al., 2008). A recent study investigating the FNR binding sites in *E. coli* by ChIP-chip and ChIP-seq showed that NAPs, such as H-NS, can often mask TF binding sites. This effect can limit the access of TFs to DNA binding sites; but it can also lead to technical artefacts in the ChIP procedure, such as, changes in the efficiency of cross-linking and immunoprecipitation (Myers *et al.*, 2013). This may explain why PhoP appears to be more strongly associated with the *mgtA* promoter than with the H-NS bound promoters. To investigate this effect further, the FLAG-tagged *phoP* gene was transduced into the *hns*-1::*kan* mutant background and the PhoP-FLAG ChIP assay was repeated in the absence of a functional H-NS protein.

However, the results of this experiment are inconclusive. Figure 5.8 A shows experimental (FLAG) and mock ChIP DNA, quantified by qPCR, normalised to Input DNA (IP/Input). Experimental ChIP DNA from all promoters is over 20-fold enriched compared to the mock ChIP DNA. Therefore, there is also a greater enrichment of the negative control, rpoD, in the experimental ChIP DNA, compared to the mock ChIP DNA. Subtraction of the mock ChIP DNA from the experimental ChIP DNA still results in enrichment of the rpoD promoter for PhoP binding (Figure 5.8 B). Thus, internal normalisation of the data to the rpoD promoter abrogates enrichment of the ugtL and STnc1480 promoters for PhoP binding remains approximately the same as in the wild-type background (Figure 5.8 C + Figure 5.7 E).

The mgtA and rpoD promoters can, therefore, no longer be used as positive and negative control promoters in this case, thus it is unclear if the assay is feasible in the hns mutant background. Inactivation of the DNA binding functions of H-NS is likely to affect normal cellular functions and pathways in S. Typhimurium as a result of altered DNA topology and overall altered transcription levels. In particular, as H-NS has been shown to bind and silence transcription from the phoP promoter (Kong et al., 2008), the absence of H-NS function is likely to result in large scale changes to the PhoP regulon. Expression of the PhoP-FLAG protein in wild type and the hns-1::kan mutant, in low Mg<sup>2+</sup> conditions, was investigated by western blot (as described in section 2.7.5). There is no difference in PhoP protein expression between wild-type and hns-1::kan strains (Figure 5.8 D), however, changes in PhoP protein activity or changes in the phosphorylation state of the PhoP protein in the absence of a functional H-NS protein cannot be ruled out. A second possibility is that the inactivation of H-NS function, and resultant changes in DNA topology, led to alterations which affected elements of the ChIP assay, such as the cross-linking efficiency. The ChIP assay may require further optimisation in a hns mutant background.

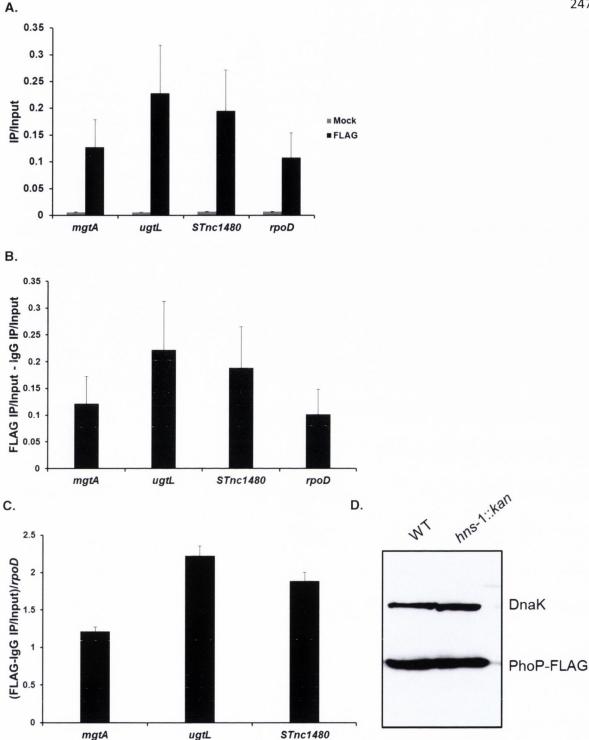


Figure 5.8 Investigation of PhoP binding to the STnc1480 promoter in an hns mutant background

A. ChIP assay provides inconclusive information about binding of PhoP to the STnc1480 promoter in the hns-1::kan phoP-3xFLAG strain. qPCR data showing ChIP DNA from the STnc1480 promoter, the positive control promoters (mgtA and ugtL) and the negative control promoter (rpoD), normalised to Input DNA. B. qPCR data of ChIP assay from two independent experiments shows experimental ChIP DNA following subtraction of mock ChIP DNA. C. qPCR data of ChIP assay from two independent experiments showing DNA from the STnc1480 and positive control promoters, internally normalised to the negative *rpoD* promoter. Error bars are based on the standard deviation from 2 independent experiments. **D.** Western blot showing expression PhoP-3xFLAG protein in wild-type and hns-1::kan strains. DnaK was probed as a loading control, as previously described (2.7.5).

# 5.2.9 Binding of SlyA to the STnc1480 promoter

The results presented so far suggest that both PhoP and SlyA are necessary for transcription from the STnc1480 promoter and that, while PhoP is necessary for transcription initiation at the STnc1480 promoter, the role of SlyA appears to be limited to the counter-silencing of H-NS. However, SlyA can contribute to a feed-forward loop, whereby SlyA binds to the *phoP* promoter and activates *phoP* transcription, likely through antagonising H-NS (Song et al., 2008). It, therefore, remains a possibility that the reduction of STnc1480 expression in the  $\Delta slyA$  mutant may reflect a reduction in *phoP* expression.

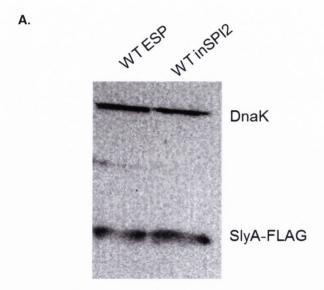
Three FLAG tags were added to the *slyA* gene to generate a C-terminally tagged SlyA protein (JH3776) (as described in section 2.5.2). Expression of SlyA-3xFLAG was investigated by western blot under two environmental conditions (ESP and inSPI2). SlyA-3xFLAG is expressed under both conditions (Figure 5.9 A). To investigate the functionality of the tagged SlyA protein, *slyA*-3xFLAG was grown under SPI2-inducing conditions and total RNA was extracted and probed for STnc1480 expression by northern blot. As shown in Figure 5.9 B, STnc1480 expression is significantly reduced in the *slyA*-3xFLAG strain, almost to the level of STnc1480 expression in the  $\Delta slyA$  mutant. These results indicate that tagging of the *slyA* gene does not affect the expression of SlyA but the tag affects the activity of the SlyA protein. The *slyA*-3xFLAG strain was used in a ChIP assay but there was no enrichment for SlyA binding in the experimental ChIP sample compared to the mock ChIP samples and compared to the negative control region (*hemX*), at published SlyA-bound promoters, such as *pagC* (data not shown).

Because the epitope tag-based approach failed to produce a functional tagged SlyA protein for use in ChIP, it was decided to carry out the ChIP assay using an antibody directed towards the native SlyA protein, to avoid affecting SlyA function. A custom polyclonal antibody directed against the SlyA peptide CASDRRAKRIKLTEKA (corresponding to nucleotides 246 to 294 of the *slyA* gene) was generated by Eurogentec (Belgium). Briefly, the synthesised peptide was coupled to the immunogenic carrier protein Keyhole Limpet Hemocyanin (KLH) and one rabbit was immunised. The serum from the final bleed of the immunised rabbit was purified by affinity chromatography using the synthesised peptide. The purified antibody and sera from pre-immune and final bleeds were tested for specificity to the synthesised peptide by indirect ELISA

(performed by Eurogentec). However, western blotting showed that the polyclonal anti-SlyA peptide antibody was unable to detect endogenous levels of SlyA protein in *S*. Typhimurium whole cell lysates from a variety of growth phases and growth conditions, following titration of the antibody and titration of the amount of cells used for generation of whole cell lysates (data not shown). The custom antibody was also unable to detect exogenous levels of SlyA protein over-expressed from the pBAD vector following induction of the P<sub>BAD</sub> promoter using 0.2% L-arabinose for 4 hours (data not shown). DnaK was used as a loading control on western blots to ensure the generation of whole cell lysates and western blotting procedure had worked correctly.

The antibody was also used in a ChIP assay using wild-type cells with the  $\Delta slyA$  mutant, as a negative control. The custom antibody did not detect SlyA *in vivo* and no specific enrichment for SlyA binding to positive control regions, such as the ugtL promoter, was detected in the wild-type strain compared to the  $\Delta slyA$  mutant (data not shown).

Binding of SlyA to the STnc1480 promoter and direct regulation of expression from the STnc1480 promoter by SlyA has, therefore, not been confirmed in this study. Indirect regulation of STnc1480 expression via SlyA-mediated control of phoP expression cannot be ruled out. However, there are some lines of evidence which suggest that the STnc1480 requirement for SlyA for expression is not mediated via PhoP. Expression of phoP is reduced approximately 4-fold in the absence of SlyA, >2-fold in the absence of OmpR/EnvZ and approximately 3-fold in the absence of SsrA/B (Appendix II). However STnc1480 expression is only reduced 7-fold in the *ompR/envZ* and *ssrA/B* mutants, while STnc1480 expression is approximately 60-fold reduced in the *slyA* mutant. The presence of PhoP, albeit at a reduced level, in the  $\Delta slyA$  mutant should still result in induction of some level of expression from the STnc1480 promoter, as is the case in the  $\triangle ompR/envZ$ and  $\triangle ssrA/B$  backgrounds. Furthermore, if the effect of the slyA deletion on STnc1480 was due to a reduction in expression of PhoP, over-expression of PhoP should rescue STnc1480 expression in the absence of SlyA. However, ectopic over-expression of PhoP does not rescue STnc1480 expression in a  $\triangle slyA$  background (Figure 5.4 A). Finally, transcription from the STnc1480 promoter does not require SlyA in the absence of a functional H-NS protein, strengthening our hypothesis that SlyA acts to antagonise H-NS mediated silencing of the STnc1480 promoter as has been demonstrated at the pagC and ugtL promoters (Perez et al., 2008). Further experimentation will be required to confirm SlyA binding to the STnc1480 promoter, however.



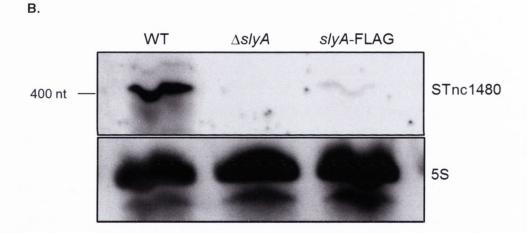


Figure 5.9 FLAG tagging the slyA gene

**A.** Western blot showing that SlyA-3xFLAG is expressed at ESP and inSPI2. DnaK was probed as a loading control. **B.** Northern blot showing that STnc1480 expression in the *slyA*-3xFLAG strain is reduced to levels resembling STnc1480 expression in the absence of SlyA, indicating that activity of the SlyA protein is affected by addition of the FLAG tags. 5S RNA was probed as a loading control.

#### 5.2.10 The role of STnc1480

STnc1480 is most highly expressed under conditions which mimic the intracellular environment, and the STnc1480 gene bears the hallmarks of having been horizontally acquired (sections 3.2.11.2 and 5.2.5). We speculate that, due to the silencing of STnc1480 expression by H-NS, and the proposed counter-silencing mechanism mediated by PhoP and SlyA, STnc1480 function has been incorporated into the existing cellular networks to provide a fitness advantage to *S*. Typhimurium under intracellular conditions. As shown in Table 3.2, a TraDIS experiment showed that a transposon inserted into the STnc1480 gene (at nucleotide 203) resulted in complete attenuation of fitness of the mutant strain following oral infection of chicken, pig and calf models. The STnc1480 transposon mutant was assigned an arbitrary fitness score of -15 as no reads corresponding to the transposon insertion site were obtained during the subsequent RNA-seq analysis of mutants in the output pool (Chaudhuri et al., 2013). These data indicate that STnc1480 may play an important role during the *S*. Typhimurium infection process.

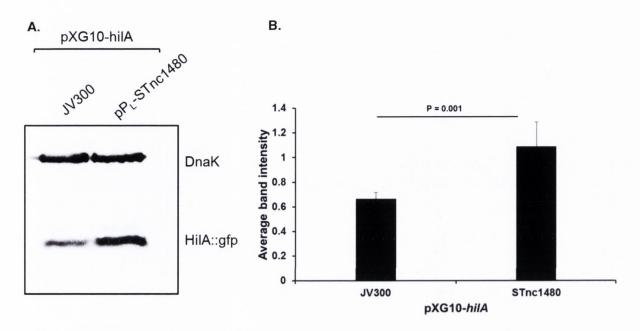
As part of a large scale study to identify the mRNA targets of *S.* Typhimurium sRNAs, the central regulator of SPI1 gene expression, *hilA*, was identified as a target for STnc1480 (Händler, 2014). The two-plasmid validation system was used, as previously described (section 2.9), to investigate if constitutive co-expression of STnc1480 affects accumulation of GFP fused to the 5'UTR and 5' end of the *hilA* gene in a heterologous system. Figure 5.10 A is a representative western blot of 3 independent biological and 9 technical replicates and demonstrates that GFP from the *hilA*::GFP fusion accumulates to a higher level when co-expressed with STnc1480 from the P<sub>LlacO</sub> promoter, compared to co-expression of the *hilA*::GFP fusion with a nonsense RNA expressed from the JV300 control plasmid (Table 2.2). Densitometry analysis of the GFP signal from the western blot in Figure 5.10A, normalised to the loading control (DnaK) signal, shows that GFP accumulation is increased approximately 1.6-fold (p = 0.001) when the pXG10-*hilA* fusion is co-expressed with pP<sub>L</sub>-STnc1480 compared to the control plasmid JV300 (Figure 5.10 B). These results indicate that STnc1480 is a direct positive regulator of *hilA* expression.

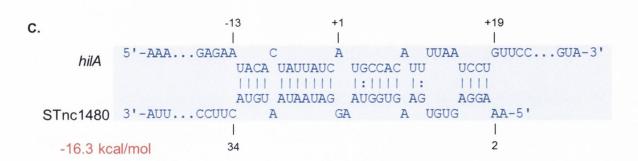
The putative interaction between STnc1480 and *hilA* was predicted using the IntaRNA software (Busch et al., 2008). As shown in Figure 5.10 C, STnc1480 putatively binds to a

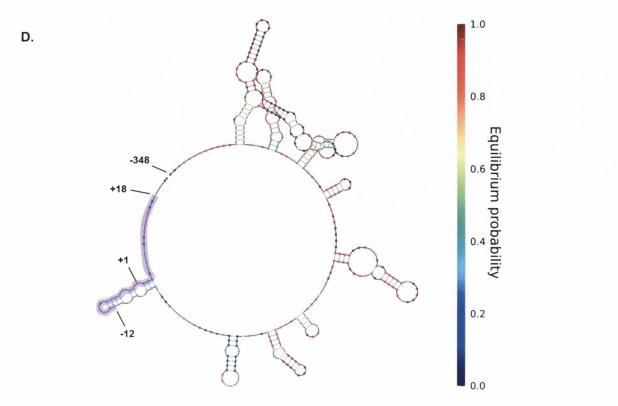
site which overlaps the initiation codon of the *hilA* mRNA, just downstream of the Shine-Dalgarno site. Nucleotides +2 to +34 (relative to the TSS) of STnc1480 are predicted to be involved in this putative interaction. The putative interacting region of the STnc1480 gene is predicted to contain a number of loop and bulge regions (Figure 5.2 B), and is thus likely to be a quite unstable region of the sRNA, as well as being within the most highly conserved region of STnc1480 (Figure 5.2 A). The predicted combined energy score necessary to make the sRNA and target binding sites accessible, as well as the energy used to hybridise both RNA molecules is -16.3 kcal/mol.

The structure of the *hilA* 5' UTR as predicted by <a href="www.nupack.org">www.nupack.org</a> is presented in Figure 5.10 D (Zadeh et al., 2011). Numbering for *hilA* mRNA is relative to the first nt of the translation initiation codon (A is +1). The *hilA* TSS is located 348 nt upstream of the first nt of the translation initiation codon (Kröger et al., 2013). These positions are indicated on the predicted structure (Figure 5.10 D). The putative STnc1480 interaction site extends from the -12 to +18 nucleotides (relative to the translation initiation codon) of the *hilA* mRNA and this region is indicated on the predicted structure. The predicted interaction region of the *hilA* mRNA incorporates a stem-loop structure which is likely to block binding of the ribosome to the RBS or to inhibit efficient translation initiation. As has been previously shown for sRNAs that activate their targets, such as DsrA activation of *rpoS* translation (Majdalani et al., 1998), binding of the sRNA at this site is likely to result in a re-structuring event that allows ribosome access to the RBS or improves the efficiency of translation (Frohlich & Vogel, 2009).

Figure 5.10 The hilA mRNA 5' UTR is a STnc1480 target







Free energy of secondary structure: -56.30 kcal/mol

#### Figure 5.10 The hild mRNA 5' UTR is a STnc1480 target

A. Western blot showing two-plasmid validation of STnc1480 activation of HilA (section 2.9). GFP accumulates to a higher level when a plasmid-borne hilA::gfp translational fusion (pXG10-hilA) is constitutively co-expressed with plasmid-borne STnc1480 (pP<sub>L</sub>-STnc1480), than with a control plasmid expressing a nonsense RNA (JV300). DnaK was probed as a loading control. **B.** Densitometry measurement using ImageJ software, based on the intensity of each western blot band, corresponding to GFP accumulation following co-expression of the pXG10-hilA plasmid with JV300 or with pP<sub>L</sub>-STnc1480, normalised to the signal intensity of the loading control, DnaK. Error bars are based on the standard deviation of five technical replicates from two independent biological replicates. C. IntaRNA prediction of the interaction between STnc1480 and hila mRNA. Numbers indicate the nucleotide position relative to the first nt of the translation initiation codon of hilA or the TSS of STnc1480. STnc1480 putatively binds in a region incorporating the hilA start codon. **D.** Predicted structure of the hilA 5' UTR and the first 7 codons of the hila mRNA using software available from www.nupack.org. putative interaction site between STnc1480 and the hilA mRNA incorporates a predicted stem-loop structure. The region predicted to interact with STnc1480 is highlighted in purple. Numbers indicate the nt position, as described in (C).

# 5.2.11 STnc1480 regulation of hilA expression

HilA is the primary regulator of SPI1 gene expression and of SPI1-mediated host intestinal cell invasion (Ellermeier & Slauch, 2007). As such, HilA expression is primarily induced in the extracellular environment of the lumen of the small intestine, by environmental signals present at this site, such as low oxygen, high osmolarity and short chain fatty acids (Lucas et al., 2000, Lawhon et al., 2002). Intracellular signals, such as low Mg<sup>2+</sup> and phosphate starvation are known to repress HilA expression via the PhoP/Q and PhoB/R TCSs, respectively (Bajaj et al., 1996, Lucas et al., 2000). Thus, activation of hilA expression by the PhoP-dependent STnc1480 is unexpected as STnc1480 is most highly induced under hilA-repressing or SPI2-inducing conditions, which mimic the intracellular lifestyle of S. Typhimurium. Figure 5.11 A shows the expression pattern of both the hilA and STnc1480 genes across a compendium of 20 environmental conditions (Kröger et al., 2013) and following growth in murine macrophages (Srikumar et al., 2014). During growth in Lennox broth, transcription from the STnc1480 promoter is induced in late exponential phase (LEP) and peaks at ESP, resembling the induction of hilA and other SPI1 genes. It was, therefore decided to investigate STnc1480 regulation of hilA expression under SPI1-inducing conditions in Lennox broth.

The JV300 control plasmid and the pP<sub>L</sub>-STnc1480 plasmid were transformed into wild-type 4/74 or  $\Delta$ STnc1480 strains resulting in the constitutive expression of a nonsense RNA and full-length STnc1480, respectively, in native *S.* Typhimurium backgrounds. Wild-type carrying the JV300 plasmid and the  $\Delta$ STnc1480 mutant carrying the JV300 or pP<sub>L</sub>-STnc1480 plasmids were grown in Lennox broth to OD<sub>600</sub>1.0 (LEP) and OD<sub>600</sub>2.0 (ESP), transcription was stopped and cells were harvested for RNA extraction. The RNA was converted to cDNA and qPCR was performed, as previously described (sections 2.6.3 and 2.6.4). Expression of *hilA* was internally normalised to the control gene *hemX*. Data from 3 independent biological experiments demonstrate no significant change in *hilA* expression between wild-type and  $\Delta$ STnc1480 at either growth condition. There is also no significant difference in *hilA* expression between the  $\Delta$ STnc1480 mutant or wild-type and the complemented mutant ( $\Delta$ STnc1480 pP<sub>L</sub>-STnc1480) at either growth phase (Figure 5.11 B).

As STnc1480 putatively binds hilA around the translation initiation codon (Figure 5.10 B and C), it is proposed that STnc1480 affects the translation efficiency of the hild mRNA. Three FLAG tags were added to the hilA gene to produce a C-terminally tagged HilA protein (strain JH3780). Wild-type and ΔSTnc1480 carrying the JV300 or pP<sub>L</sub>-STnc1480 plasmids (and FLAG-tagged HilA) were grown in Lennox broth, and cells were harvested for analysis by western blot at LEP and ESP (Figure 5.11 C). At LEP there is a decrease in HilA expression in the ΔSTnc1480 JV300 and ΔSTnc1480 pP<sub>L</sub>-STnc1480 strains compared to wild-type. At ESP, when HilA expression is at its peak, there is very little difference in HilA expression between wild-type and  $\Delta STnc1480$  JV300. However, constitutive over-expression of STnc1480 in the ΔSTnc1480 pP<sub>L</sub>-STnc1480 strain results in a large increase in HilA expression compared to wild-type at ESP (Figure 5.11 C). Densitometry analysis of the western blot bands, from Figure 5.11 C, is presented in Figure 5.11 D. In ΔSTnc1480 JV300 and ΔSTnc1480 pP<sub>L</sub>-STnc1480 at LEP, HilA expression is approximately 1.6-fold reduced, reflecting the reduction in HilA::GFP expression, presented in Figure 5.10 B. At ESP, HilA expression in the ΔSTnc1480 pP<sub>L</sub>-STnc1480 strain increases approximately 1.7-fold, compared to wild-type, and 1.4-fold, compared to  $\Delta$ STnc1480 JV300.

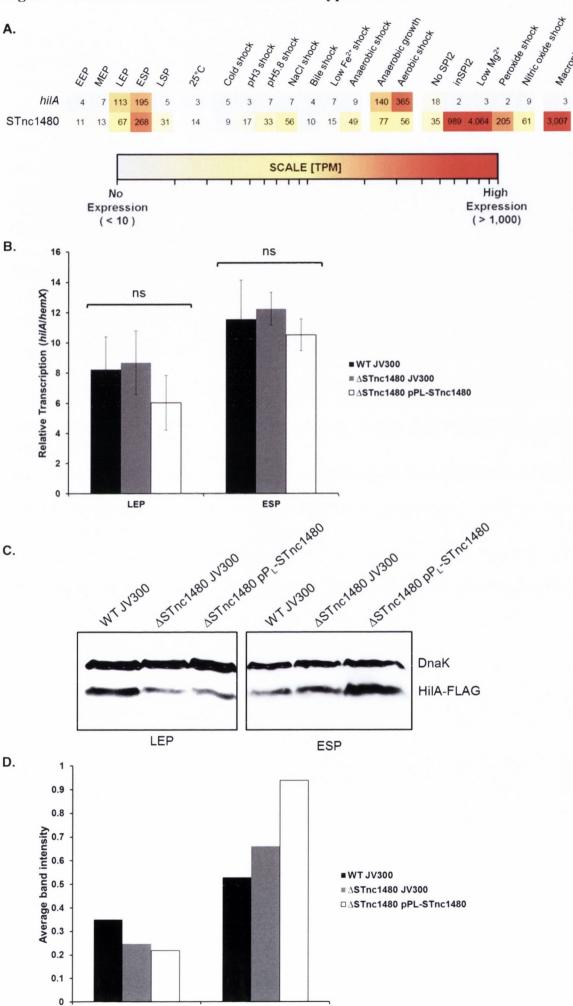
These data suggest that STnc1480 acts by targeting the translation efficiency of hilA mRNA. The lack of difference in hilA expression levels between wild-type and  $\Delta$ STnc1480, as determined by qPCR, at LEP and ESP may suggest that STnc1480 does

not affect *hilA* mRNA steady state levels. In the previously mentioned, well-studied case of DsrA activation of *rpoS* translation, *rpoS* mRNA is rapidly degraded by RNase III in the absence of DsrA (Resch *et al.*, 2008). However, without further experimentation, we cannot distinguish between transcription and turnover of the *hilA* mRNA, and cannot speculate if STnc1480 binding also protects *hilA* mRNA from cleavage.

The most important time for STnc1480-mediated activation of HilA expression is at the end of exponential growth or the entry into stationary phase, when invasion-associated genes are beginning to be induced. At LEP differential expression of HilA between the wild-type strain and ΔSTnc1480 is clear. It is unclear why constitutive over-expression of STnc1480 from the pP<sub>L</sub>-STnc1480 plasmid does not restore HilA expression to wild-type levels at LEP, however. At later timepoints during growth under invasion-inducing conditions, hild expression is under the direct or indirect positive control of other factors, such as the HilD-HilC-RtsA feed-forward loop (Ellermeier et al., 2005), Fur (Troxell et al., 2011b), BarA/SirA and the Csr system (Fortune et al., 2006), FliZ (Chubiz et al., 2010) and Dam (Balbontin et al., 2006). The apparent functional redundancy between STnc1480 and these positively acting regulatory systems may mask the effects of the absence of STnc1480, thus there is no change in HilA expression in the ΔSTnc1480 mutant at ESP. However over-expression of STnc1480 at ESP does lead to an accumulation of HilA, perhaps due to the action of the previously mentioned positive regulators of hilA transcription, combined with STnc1480-mediated translational activation of HilA.

To definitively confirm the interaction between STnc1480 and the *hilA* mRNA 5' UTR, however, single nucleotide exchanges in the putative interaction site of STnc1480 or *hilA*, followed by compensatory mutations, are required. These experiments will directly validate the GFP-based results presented in Figure 5.10 A and the predictions presented in Figure 5.10 C and D, as has been previously demonstrated for SgrS targeting of *sopD* (Papenfort et al., 2012).

Figure 5.11 STnc1480 activation of HilA in S. Typhimurium



ESP

LEP

#### Figure 5.11 STnc1480 activation of HilA in S. Typhimurium

**A.** Screenshot from SalCom website showing expression of STnc1480 and *hilA* across 20 environmental conditions (Kröger et al., 2013) and within murine macrophages (Srikumar et al., 2014). In Lennox medium, expression of STnc1480 and *hilA* increases at LEP and ESP. **B.** qPCR data from three independent biological replicates, showing no significant change in *hilA* expression levels between wild-type JV300, ΔSTnc1480 JV300 and ΔSTnc1480 pP<sub>L</sub>-STnc1480 at LEP and ESP. *hilA* transcription was quantified relative to a standard curve of genomic DNA and internally normalised to the *hemX* gene, as previously described (section 2.6.4). Error bars are based on the standard error from 3 independent biological replicates. **C.** Western blot showing expression of HilA-3xFLAG protein in the same strains as panel (**B**) grown to LEP and ESP. DnaK was used as a loading control. **D.** Densitometry measurement, using ImageJ software, based on the intensity of each western blot band from panel (**C**), normalised to the signal intensity of the loading control, DnaK.

# 5.2.12 The effect of STnc1480 on Type Three Secretion systems

Analysis of the secreted proteins within culture supernatants from S. Typhimurium grown under SPI1-inducing conditions provides a rough measure of the intactness of the SPI1 TTSS and of SPI1 gene expression (Pfeiffer et al., 2007). To investigate if removal of the gene encoding STnc1480, and concomitant reduction in HilA expression, affects the formation of a functional SPI1 TTSS, culture supernatants from wild-type and  $\Delta$ STnc1480 independently grown to  $OD_{600}0.4$  and  $OD_{600}1.6$  were analysed by SDS-PAGE, as previously described (section 2.7.2) . There is no apparent change in the secretion profile of SPI1 effectors between wild-type and  $\Delta$ STnc1480 at either growth phase (Figure 5.12). Early during exponential growth, most SPI1 effectors are not expressed or secreted. At later timepoints, as SPI1 secretion is induced, we speculate that the effects of the STnc1480 mutation are masked by other positive regulators of *hilA* expression, resulting in the maintenance of a fully functional TTSS.

However, unlike the SPI1 TTSS, the flagellar TTSS appears to be affected by deletion of the STnc1480 gene. Production or secretion of the flagellin subunit protein FliC is increased in the  $\Delta$ STnc1480 mutant compared to wild-type at both growth phases (Figure 5.12 A and B), suggesting that STnc1480 represses expression or secretion of FliC. A 10 minute induction of STnc1480 over-expression with 0.2% L-arabinose, from the  $P_{BAD}$  promoter, in a wild-type background leads to a decrease in the level of FliC expression or secretion. Addition of L-arabinose to the wild-type strain carrying an empty pBAD

vector does not result in a decrease in the level of FliC expression or secretion (Figure 5.12 C). The presence of a slightly higher molecular weight band than FliC in the wild-type pBAD-STnc1480 strain, may indicate a switch towards the alternative flagellin subunit FljB, which is slightly larger than FliC (Schmitt *et al.*, 1996). Use of the two-plasmid validation system to investigate if STnc1480 interacts with a region within the 5' UTR of *fliC* mRNA did not result in differential GFP accumulation (data not shown). It is, thus, unclear if STnc1480 directly affects expression of FliC or if STnc1480 affects the phase variation of the genes encoding the flagellin subunits.

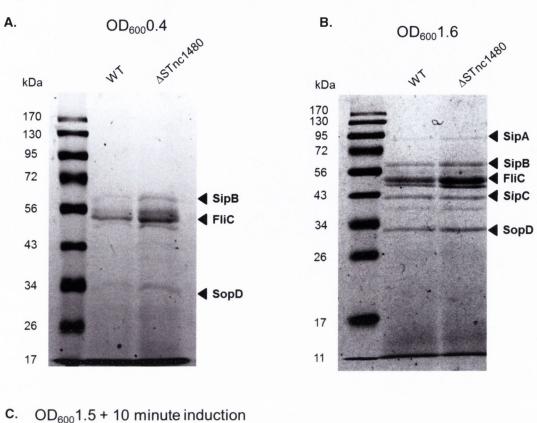
We attempted to use western blotting to differentiate between the alternate flagellin subunits, to make use of a loading control, allowing more accurate determination of the differential expression of FliC in wild-type and  $\Delta$ STnc1480 strains, and to probe for levels of FliC protein within whole cell lysates to investigate if altered levels of secreted FliC protein in  $\Delta$ STnc1480 were as a result of changes in FliC expression or FliC secretion. However, transduction of a *fliC*-3xFLAG gene (Hautefort et al., 2008) into the 4/74 background resulted in over-expression of the FliC protein and a non-motile phenotype in a wild-type background. The loss in motility and increased FliC expression did not occur in the  $\Delta$ STnc1480 background (Figure 5.12 D and E). It is unclear why tagging of the 4/74 FliC protein in a wild-type background led to loss in motility and over-expression of FliC, as the original SL1344 strain remained motile when the *fliC* gene was tagged (Hautefort et al., 2008). Therefore, we were unable to characterise the molecular interaction between STnc1480 and FliC any further.

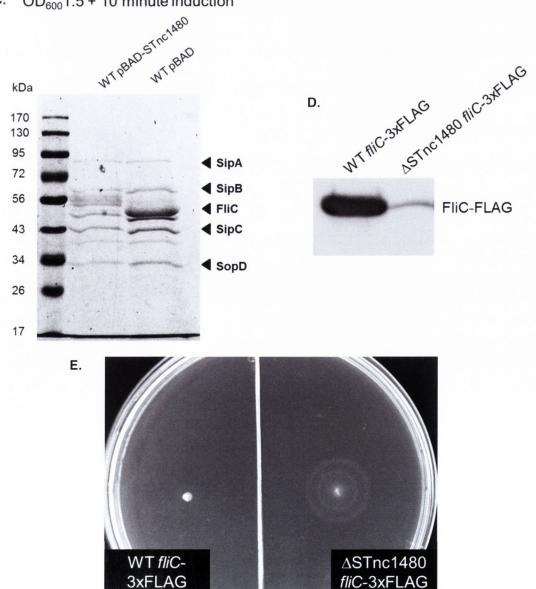
Despite the apparent negative regulation of FliC expression or secretion by STnc1480, the  $\Delta$ STnc1480 mutant does not exhibit a differential swimming phenotype when assayed for motility, as previously described (section 2.2.4). Figure 5.12 F is representative of the measurement of bacterial swimming diameters of 6 technical replicates and 2 biological replicates (replicates graphed in Figure 5.12 G) and shows that there is no significant difference in the swimming ability of  $\Delta$ STnc1480, compared to wild-type. However, previous experiments from our laboratory showed that STnc1480 constitutively over-expressed from the P<sub>LlacO</sub> promoter does lead to a significantly reduced swimming phenotype compared to the wild-type strain expressing a nonsense RNA from the P<sub>LlacO</sub> promoter. The assay was performed in triplicate and significance was considered to be 3 times larger than the standard deviation of the wild-type diameters. This effect was shown to be independent of growth rate of the  $\Delta$ STnc1480 mutant (Händler, 2014).

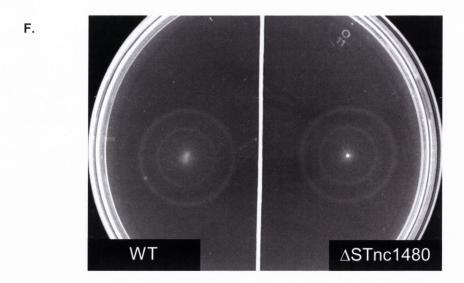
These data show that increased expression or secretion of FliC in the ΔSTnc1480 mutant strain does not affect motility, but the reduction in FliC caused by over-expression of STnc1480 does inhibit the swimming capabilities of the bacteria. We may speculate, from these results, that STnc1480 does not merely affect the phase variable switching of the flagellin subunit from FliC to FljB. Flagellar phase variation occurs by a site-specific DNA inversion event. The promoter of the gene encoding the alternate flagellin subunit, FljB, and the downstream gene encoding the regulatory protein, FljA, is flanked by recombination sites hixR and hixL. The recombinase, Hin, and recombination enhancer proteins, Fis and HU, mediate reversible recombination between the hix sites, resulting in inversion of the region of DNA containing the fljBA promoter. In one orientation transcription from the fljBA promoter is switched on and fljBA is transcribed, resulting in FliB being secreted as the flagellin subunit. FliA acts to repress expression of the unlinked fliC gene. Inversion of the fljBA promoter in the alternate orientation results in the switching-off of fljBA transcription. In the absence of FljA, repression of fliC is removed and FliC is secreted as the flagellin subunit (Bonifield & Hughes, 2003). Increased levels of FljB, rather than FliC, would not be expected to result in a motility defect as S. enterica typically switches between expression of FliC and FljB at a rate of 10<sup>-3</sup> to 10<sup>-5</sup> per cell generation (Bonifield & Hughes, 2003), thus, we speculate that STnc1480 may be involved in regulating the expression of FliC.

A recent study to identify novel genes involved in regulation of the motility of *S*. Typhimurium has highlighted the complexities and unexplained regulatory effects involved in flagellar gene regulation (Bogomolnaya *et al.*, 2014). Further investigation is necessary to elucidate the mechanisms and factors involved in the putative interaction between STnc1480 and FliC expression or secretion.

Figure 5.12 The effect of STnc1480 on secreted levels of FliC







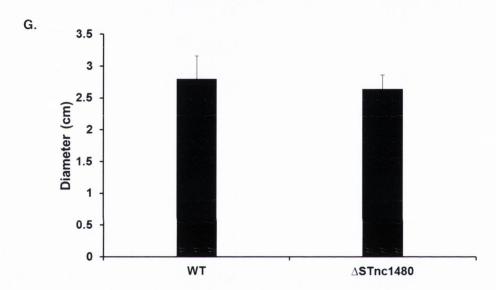


Figure 5.12 The effect of STnc1480 secreted levels of FliC

A. The secreted protein profile of wild-type and  $\Delta STnc1480$  mutant grown to  $OD_{600}0.4$ and **B.** OD<sub>600</sub>1.6 was analysed by SDS-PAGE and Coomassie staining, as previously described (sections 2.7.2, 2.7.3 and 2.7.4). The SPI1 effector protein secretion pattern is not altered in the \Delta STnc1480 mutant but expression of the flagellin protein FliC is increased in both growth phases (independent cultures) in the ΔSTnc1480 mutant. C. The secreted protein profile of wild-type strains carrying the pBAD-STnc1480 plasmid and the empty pBAD plasmid. Both strains were grown to OD<sub>600</sub>1.5, and expression from the pBAD vector was induced using 0.2% L-arabinose for 10 minutes. Expression of the flagellin protein FliC is reduced upon over-expression of STnc1480. **D.** Western blot of secreted FliC-3xFLAG protein, showing over-expression of FliC-3xFLAG in the wild-type background, compared to the ΔSTnc1480 background. E. Motility assay plate demonstrating that FLAG-tagging fliC in a wild-type background results in a non-motile phenotype. ΔSTnc1480 fliC-3xFLAG remains motile. F. Representative motility assay plate showing no difference in the ability of a ΔSTnc1480 mutant to swim compared to wild-type. G. Bar chart shows the swimming diameters of wild-type and ΔSTnc1480. Error bars are based on the standard deviation from 2 independent biological replicates and 6 technical replicates.

#### 5.3 Discussion

# 5.3.1 Transcriptional regulation of STnc1480

STnc1480 is a large intergenic sRNA that was classified as a SPI2-like sRNA based on its pattern of regulation (this study) and on its pattern of expression across a compendium of 22 environmental conditions (Kröger et al., 2013). STnc1480 expression is also highly induced within murine macrophages, reflecting the resemblance to SPI2 genes (Srikumar However, further investigation of the transcriptional regulation of et al., 2014). STnc1480 indicated that STnc1480 is regulated and expressed more similarly to genes, such as pagC, ugtL and mig-14 that require both PhoP and SlyA for their expression. Expression of STnc1480 is approximately 60-fold decreased in  $\Delta phoP$  and  $\Delta slyA$ mutants. STnc1480 expression is highest under low Mg<sup>2+</sup> conditions, a signal that is sensed by PhoQ, the membrane bound sensor kinase of the PhoP/Q TCS, leading to an induction of expression of genes in the PhoP regulon (Garcia Vescovi et al., 1996). Additionally, ectopic expression of PhoP and SlyA can restore STnc1480 expression in PhoP and SlyA isogenic mutant backgrounds, respectively, but ectopic expression of neither protein can restore STnc1480 expression in the absence of the other protein. Expression from the STnc1480 promoter is fully dependent on the presence of both PhoP and SlyA.

It has previously been shown that a subset of PhoP-dependent genes also require SlyA for their expression (Navarre et al., 2005) and this was demonstrated to be a mechanism of counteracting the transcriptional silencing effect of H-NS (Perez et al., 2008). STnc1480 is not well conserved outside of the *Salmonella* genus and the full STnc1480 sequence is not present outside of certain serovars of *Salmonella enterica*. The STnc1480 gene bears the hallmarks of having been acquired through lateral gene transfer due to its high AT content and its low level of phylogenetic distribution, suggesting that the gene was acquired recently in *Salmonella* evolution. The STnc1480 promoter is bound by H-NS and expression from the STnc1480 promoter is silenced by H-NS under conditions which do not induce STnc1480 expression (MEP). H-NS remains associated with the STnc1480 promoter under conditions when STnc1480 is highly induced suggesting silencing of the STnc1480 promoter is relieved through promoter re-structuring, rather than the displacement of H-NS from the promoter that has been demonstrated for SlyA-mediated counter-silencing of the *hlyE* promoter in *E. coli* (Lithgow et al., 2007).

In the absence of a functional H-NS protein, STnc1480 no longer requires SlyA for transcription, while STnc1480 transcription still requires PhoP. This indicates that SlyA is necessary only to counter-silence the repressive effect of H-NS, while PhoP is necessary for RNAP recruitment and for initiation of STnc1480 transcription. PhoP directly binds to and regulates expression from the STnc1480 promoter. However, SlyA binding to the STnc1480 has not been demonstrated. It, therefore, remains a possibility that the decrease in STnc1480 expression in the  $\Delta slyA$  mutant is an indirect effect as phoP expression is positively regulated by SlyA (Song et al., 2008). However, as previously discussed and based on the fact that STnc1480 is expressed in a  $\Delta slyA$  mutant in a hns mutant background, the counter-silencing model is a likely mechanism of STnc1480 transcriptional regulation.

#### 5.3.2 STnc1480 and S. Typhimurium virulence

The apparent horizontal acquisition of the STnc1480 gene and the subsequent counter-silencing of STnc1480 expression under conditions which mimic the host intracellular environment suggest that STnc1480 gene function has provided *S*. Typhimurium with a fitness benefit under these conditions. The complete attenuation of fitness of a mutant carrying a transposon insertion in the STnc1480 gene during infection of food-producing animals (Chaudhuri et al., 2013) strengthens the hypothesis that STnc1480 is an important sRNA during the infection process, either by directly affecting virulence genes or by affecting survival of *S*. Typhimurium during infection.

The discovery that STnc1480 directly targets and up-regulates expression of the central SPI1 regulator HilA, was surprising. STnc1480 apparently binds the *hilA* mRNA in a region downstream of the Shine-Dalgarno site but overlapping the translation initiation codon and extending into the first 6 codons of the *hilA* mRNA. Binding in this region is common for *trans*-acting sRNAs that activate their targets, as target activation by sRNAs typically involves re-structuring of secondary structures that limit the translation efficiency of target mRNA (Frohlich & Vogel, 2009). Software available from <a href="https://www.nupack.org">www.nupack.org</a> predicts the presence of a stem-loop structure in the *hilA* mRNA, at a site which overlaps the putative STnc1480 interaction site. Rearrangement of this stem-loop structure due to STnc1480 binding is likely to enhance the translation efficiency of the *hilA* mRNA. This interaction appears to occur in the later stages of

exponential growth, just before the general induction of SPI1 genes. The STnc1480-mediated activation of HilA expression during this growth phase may occur to provide basal levels of HilA protein prior to induction of the SPI1 regulatory cascade.

However the high levels of STnc1480 expression under conditions which mimic the intracellular environment, especially low Mg<sup>2+</sup> PCN, suggests that this sRNA should play an important role under these conditions. However, HilA is not expressed under macrophage-mimicking conditions (Kröger et al., 2013). It is possible that STnc1480 has further, as yet unidentified, targets that play a role during infection of the macrophage. An additional possibility is that STnc1480 activation of HilA expression occurs in the intra-epithelial cell environment. Salmonella invades epithelial cells of the host intestine and survives and replicates within Salmonella containing vacuoles (SCV) in this host cell type (Knodler & Steele-Mortimer, 2003). It has previously been shown that, unlike the macrophage environment, which specifically induces the SPI2 TTSS and down-regulates the SPI1 and flagellar TTSSs, all three TTSSs are induced in the intra-epithelial cell environment (Hautefort et al., 2008). SPI2 genes and the secretion apparatus remain highly expressed at 2, 4 and 6 hours post epithelial cell infection, while expression of the SPI1 TTSS increases with time during epithelial cell infection. The flagellar TTSS is not expressed early during infection of epithelial cells but becomes induced later during infection (Hautefort et al., 2008). The epithelial cell environment was found to contain some similar environmental signals to the macrophage environment. The levels of magnesium and phosphate ions are low in both host cell types. However, unlike in the macrophage environment, genes involved in resistance to oxidative and nitrosative stress as well as in acid resistance are not induced in the epithelial cell environment, indicating that the epithelial cell is not as bactericidal as phagocytic cells (Hautefort et al., 2008). STnc1480 involvement in gene regulation under the epithelial cell conditions fits with the expression pattern of STnc1480, which is highly induced under low Mg<sup>2+</sup> conditions but it not as highly induced in response to peroxide and nitric oxide shock as archetypical SPI2 genes are (Figure 5.3 A). It is, thus, possible that STnc1480 provides a method of inducing SPI1 gene expression in response to the epithelial cell cues.

The putative repression of FliC by STnc1480 may also be physiologically relevant under epithelial cell conditions as the flagellar apparatus is not expressed during early epithelial cell infection, and becomes induced after 4 hours within epithelial cells. Further analysis

will be required to validate this model and to assess the effect of deleting or over-expression of STnc1480 during epithelial cell infection.

Figure 5.13 is a model summarising our current knowledge and hypotheses regarding the regulation of STnc1480 transcription and how STnc1480 fits within the established regulatory networks of *S*. Typhimurium.

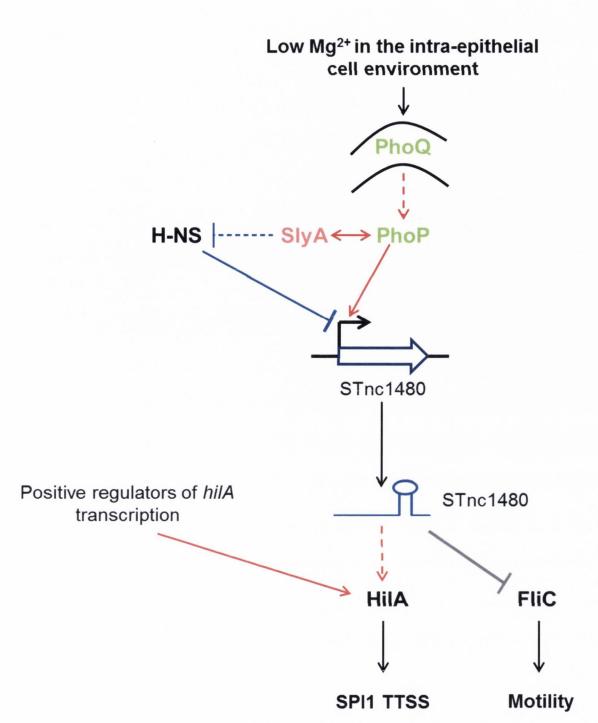


Figure 5.13 Model of the transcriptional regulation and the role of STnc1480

Schematic model summarising our current knowledge and hypotheses regarding the transcriptional regulation of STnc1480, and the role of STnc1480 within *S.* Typhimurium. STnc1480 transcription is silenced by H-NS. The silencing mechanism is counteracted by SlyA, and STnc1480 transcription is activated by PhoP. STnc1480 post-transcriptionally activates expression of the central SPI1 regulator, HilA. The action of STnc1480 complements the actions of positive regulators of *hilA* transcription. We hypothesise that this interaction occurs in the SCV within host epithelial cells. STnc1480 also directly or indirectly represses expression or secretion of the flagellin subunit, FliC, via an unknown mechanism. Full red arrows denote transcriptional activation, dotted red arrows denote post-transcriptional activation of expression or activity. Full blue T-bars denote transcriptional repression, dotted blue T-bars denote post-transcriptional repression of expression or activity. Grey T-bar denotes repression by an unknown mechanism.

# Chapter 6 General Discussion

# 6.1 Importance of this study

### 6.1.1 Context of the study

Salmonella enterica is a medically important zoonotic pathogen that causes a range of diseases from self-limiting gastroenteritis to Typhoid fever. Non-host adapted serovars of *S. enterica*, such as *S.* Typhimurium, are considered to be "successful" pathogens as these serovars can infect a wide range of hosts, survive a series of stressful events during the infection process and, in susceptible hosts, disseminate and cause systemic infection (Fabrega & Vila, 2013) (Figure 1.1).

S. Typhimurium establishes infection by integration of complex environmental signals through regulatory molecules, such as TFs,  $\sigma$ -factors and small RNAs, allowing the coordination of gene expression at the transcriptional and post-transcriptional levels. The careful coordination of gene expression by these regulatory molecules, in response to environmental, spatial and temporal cues, ensures that the bacterium expresses only the genes necessary for survival and proliferation under certain environmental conditions, while avoiding loss of bacterial fitness due to inappropriate or wasteful gene expression (Groisman & Mouslim, 2006, Sengupta et al., 2014). An understanding of the complex signals, regulatory inputs and regulatory pathways that result in the transcriptional activation or repression of virulence-associated genes provides a clearer picture of the steps involved in establishment of S. Typhimurium infection. Many studies have focussed on the signals which control expression and/or activity of TFs and σ-factors, and concomitantly of their target genes (Bajaj et al., 1996, Garcia Vescovi et al., 1996, Deiwick et al., 1999, Bang et al., 2002, Norte et al., 2003, Balbontin et al., 2006, Ellermeier & Slauch, 2007, Fass & Groisman, 2009, Yoon et al., 2009, Chubiz et al., 2010, Troxell et al., 2011b, Golubeva et al., 2012), however our knowledge of the regulatory factors involved in controlling expression of sRNA genes has lagged behind.

Typically, sRNAs have also been overlooked in large-scale investigations into the regulatory pathways involved in controlling gene expression in bacteria (Gottesman, 2005). Many sRNAs are encoded in IGRs that were originally considered to be "empty" areas of the chromosome, and the early DNA microarrays contained probes directed towards coding sequences only. The use of large-scale proteomic studies, too, provided a false view that protein-mediated gene regulation was the only method of gene regulation

within the cell. The discovery of vast numbers of sRNAs, with diverse mechanisms of action and physiological functions, in many studied bacterial species, highlights the key role played by sRNAs in bacterial gene regulation, usually at the post-transcriptional level (Vogel, 2009). sRNAs target a diverse range of genes and there have been a number of examples of sRNA-mediated regulation of expression of the genes which encode regulatory proteins, (Majdalani et al., 1998, Majdalani et al., 2001, Mandin & Gottesman, 2010, Altuvia et al., 1997, Holmqvist et al., 2012, Corcoran et al., 2012, Holmqvist et al., 2010, Coornaert et al., 2010, Guillier & Gottesman, 2008). Furthermore, the regulatory proteins involved in controlling expression of sRNAs have been elucidated in some cases (see Table 3.1) and reviewed in (Gopel & Gorke, 2012, Mandin & Guillier, 2013). These examples highlight the intimate connection between transcriptional post-transcriptional regulatory elements of the cell. To our knowledge, however, this is the first study that has aimed to characterise the extent of the interconnection between transcriptional and post-transcriptional regulators, on a global scale, through investigation of the protein factors involved in transcriptional regulation of sRNA genes.

In this study we used an RNA-seq-based transcriptomic approach to expand the regulons of key virulence-associated regulatory proteins, under infection-relevant conditions, in *S*. Typhimurium. The use of RNA-seq, rather than more traditional methods of transcriptomic analysis, has given us a view of the *S*. Typhimurium transcriptome at single nucleotide resolution and unprecedented depth (Ozsolak & Milos, 2011). RNA-seq provides an unbiased picture of all genes which are being actively transcribed, with no requirement for previous knowledge of the sequence which is being transcribed (Ozsolak & Milos, 2011, Croucher & Thomson, 2010). Thus, RNA-seq is the ideal approach to use for the investigation of the expression or regulation of sRNAs, and for the discovery of new sRNAs.

#### 6.1.2 Global analysis of the transcriptional regulation of sRNAs

Our laboratory previously identified and characterised the environmental conditions that induce expression of 280 sRNAs in S. Typhimurium (Kröger et al., 2013). By analysing the transcriptome of 16 mutants lacking important  $\sigma$ -factors, TFs and TCSs under 5 environmental conditions we have further characterised many of these sRNAs by identifying the regulatory proteins which may be involved in controlling the expression of

these sRNAs at the transcriptional level. As has been described for FnrS and RyhB, sRNAs under the control of important regulatory proteins may provide the "missing link" between the regulatory protein and genes within the regulon of that protein (Masse et al., 2005, Boysen et al., 2010, Durand & Storz, 2010).

We have identified 177 sRNA genes which are controlled by at least one of the regulatory proteins under investigation in this study (section 3.2.6). In general, the pattern of chromosomal location and requirement for binding by the RNA chaperone protein, Hfq, in the group of differentially expressed sRNAs reflects the patterns that had been previously identified for all S. Typhimurium sRNAs on a global scale (Kröger et al., 2013). This observation indicates that the ability of a regulatory protein to control transcription of an sRNA gene is not affected by the chromosomal location of the sRNA gene. A requirement for Hfq, also does not affect control of sRNA gene expression by this panel of protein factors. Generally, the size of the sRNA-based regulon of each regulatory protein mirrors the relative size of the protein-coding regulon of each regulatory protein. The exception is the  $\sigma^{38}$  regulon (Figure 3.7 D and E).  $\sigma^{38}$  directly or indirectly controls the largest number of sRNA genes, while the protein-coding  $\sigma^{38}$ regulon is relatively small. The apparent differences in the regulatory powers of  $\sigma^{38}$ , on sRNA and protein-coding genes, suggests that  $\sigma^{38}$  is a central hub for integrating post-transcriptional regulatory signals, a speculation reflected in the fact that four sRNAs are involved in controlling RpoS expression, post-transcriptionally. Thirty-two sRNA genes which are most likely to directly require  $\sigma^{38}$  for transcription, based on promoter sequence, are listed in Figure 3.9 G. A prominent role of  $\sigma^{38}$  in regulating expression from sRNA promoters is unsurprising as many sRNAs are known to be induced in response to various stresses and in stationary phase, as sRNAs can provide rapid adaptation to stressful environmental conditions (Beisel & Storz, 2010, Shimoni et al., 2007). The mirroring of the regulatory powers of the remaining TFs and  $\sigma$ -factors under investigation suggest that sRNAs and sRNA function is intimately connected with the well-established cellular regulons.

From the group of differentially expressed sRNA genes, in our panel of regulatory mutants, we have generated a transcriptional regulatory network (Figure 3.10). The regulatory network is a tool of systems biology that is used to show connected cellular and molecular processes (McDermott et al., 2009). Regulatory networks, such as that

presented in Figure 3.10, provide a more comprehensive picture of the interactions and inputs to transcriptional regulation within the cell that may not emerge through investigation of single components of the system (Golubeva et al., 2012, Hebrard et al., 2011). Analysis of the regulatory network showing transcriptional inputs to S. Typhimurium sRNAs, highlights the complexities of the transcriptional regulatory network and allows us to speculate that sRNAs may play important physiological roles in integrating regulatory signals and connecting diverse cellular regulons. MicA has already been shown to act as such a connector molecule between the important  $\sigma^E$  and PhoP regulons in E. coli, linking the low  $Mg^{2+}$ -inducible PhoP regulon to cell envelope stress (Coornaert et al., 2010).

Furthermore the use of a panel of mutants lacking important regulatory proteins, rather than just a single regulatory mutant, for the characterisation of differential sRNA expression, allows us to detect patterns of expression, and to group sRNAs into clusters that appear to be co-regulated. The "guilt by association" hypothesis posits that groups of co-regulated or co-expressed genes have similar cellular functions (Quackenbush, 2003). We adopted the "guilt by association" hypothesis and identified 4 sRNAs which display a "SPI1-like" pattern of expression and 10 sRNAs which display a "SPI2-like" pattern of expression across our panel of regulatory mutants (section 3.2.10). In general, the SPI-like sRNAs were most highly conserved within the Salmonella genus, and, in particular, the SPI2-like sRNAs were most highly conserved in S. enterica (Figure 3.12). The SPI-like sRNAs represent interesting candidates for further analysis of the role played by these sRNAs in S. Typhimurium virulence. Two of the SPI2-like sRNAs, STnc3020 and STnc3090, contain putative SsrB binding sites within their promoters. This suggestion, that SsrB directly binds and activates transcription of these two sRNAs, may indicate that STnc3020 and STnc3090 play a role in the intracellular survival and/or proliferation of S. Typhimurium. Five of the fourteen SPI-like sRNAs were attenuated for infection in the chicken, pig and calf models following insertion of a transposon into the sRNA genes (Chaudhuri et al., 2013), further highlighting an important role during the S. Typhimurium infection process. Two of the SPI-like attenuated sRNAs, STnc520 and STnc1480, were selected for detailed mechanistic analysis in this study.

#### 6.1.3 STnc520

STnc520 was characterised as a SPI1-like sRNA based on the high correlation of STnc520 expression with the expression of the archetypical SPI1 gene, *prgH*, across the panel of 18 regulatory mutants (this study) and across 22 environmental conditions (Kröger et al., 2013). STnc520 is encoded in an IGR of SPI11, and its expression is dependent on the primary SPI1 TF, HilD. However ectopic expression of HilD in the absence of the entire SPI1 locus does not restore STnc520 expression, indicating that HilD does not directly activate STnc520 expression. We have shown that the SPI1-encoded regulatory protein, SprB, binds and regulates expression from the STnc520 promoter (Figures 4.5 and 4.7). To our knowledge, the STnc520 promoter is only the third gene promoter shown to be directly regulated by SprB, along with the promoters of the *hilD* and the SPI4-encoded *siiA*.

The attenuation for infection of a mutant carrying a transposon insertion within the STnc520 gene (Chaudhuri et al., 2013) and the regulation of STnc520 gene expression by a SPI1-encoded virulence-associated regulator, in response to signals that activate expression of SPI1 genes, suggest that STnc520 plays an important role in *S*. Typhimurium virulence or survival during infection. We have not determined the direct targets of STnc520-mediated gene regulation, however, genes involved in formate metabolism, sulphur metabolism and copper transport are differentially expressed in response to deletion or over-expression of the STnc520 gene. A role for STnc520 in regulation of one or more of these cellular processes, suggests that STnc520 may be involved in protecting the cell against oxidative damage caused by toxic intermediate compounds during the intracellular lifestyle of *S*. Typhimurium. *ssaG*, the gene encoding the needle subunit of the SPI2 TTSS is also putatively activated by STnc520. However, a direct interaction between STnc520 and one or more of the putative target genes remains to be demonstrated.

#### 6.1.4 STnc1480

STnc1480 was characterised as a SPI2-like sRNA based on the high correlation of STnc1480 expression with the expression of the archetypical SPI2 gene, *ssaG*, across the panel of 18 regulatory mutants (this study) and across 22 environmental conditions and

within murine macrophages (Kröger et al., 2013, Srikumar et al., 2014). STnc1480 bears the hallmarks of having been horizontally-acquired recently in Salmonella evolution, based on the high AT content of the STnc1480 gene and the limited phylogenetic distribution of the STnc1480 gene sequence (section 5.2.5). STnc1480 is silenced by the xenogenic silencing protein H-NS (Figure 5.5) and requires both PhoP and SlyA for expression. Under low Mg<sup>2+</sup> conditions, PhoP and SlyA counter-silence the repressive effects of H-NS at the STnc1480 promoter. SlyA is not fully required for STnc1480 expression in the absence of a functional H-NS protein, indicating that SlyA functions only to counteract the effects of H-NS. H-NS remains bound at the STnc1480 promoter under low Mg<sup>2+</sup> conditions, indicating that SlyA counteracts H-NS-mediated silencing of the STnc1480 promoter through re-structuring of the STnc1480 promoter, rather than displacement of H-NS (Figure 5.6). PhoP directly binds to the STnc1480 promoter and is required for STnc1480 transcription, even in the absence of H-NS, indicating that PhoP is likely to act as a classical TF through recruitment of RNAP or binding to RNAP subunits to initiate STnc1480 transcription (Figure 5.7). To our knowledge, STnc1480 represents the first example of an S. Typhimurium sRNA that is subject to xenogenic silencing by H-NS, and subsequent counter-silencing mediated by PhoP and SlyA.

The attenuation of the mutant strain carrying a transposon insertion in the STnc1480 gene (Chaudhuri et al., 2013) suggests an important role for STnc1480 in *S.* Typhimurium pathogenesis. STnc1480 post-transcriptionally activates expression of the primary SPI1 regulator HilA. We speculate that the STnc1480-mediated activation of HilA occurs via an anti-antisense mechanism, in which STnc1480 binds and re-structures an inhibitory stem-loop structure within the 5' UTR of the *hilA* mRNA, thereby increasing the translation efficiency of the *hilA* mRNA (Figures 5.10 and 5.11). Activation of *hilA* expression by low Mg<sup>2+</sup>-induced STnc1480 is in contrast to the well-documented environmental cues that activate and repress *hilA* expression (Bajaj et al., 1996, Lucas et al., 2000), leading us to speculate that STnc1480-mediated activation of *hilA* expression may occur during intracellular infection of host epithelial cells. A role for STnc1480 during survival in the intra-epithelial cell environment remains to be confirmed, however, a key role in the survival and establishment of successful *S.* Typhimurium infection for STnc1480 is suggested by the acquisition and maintenance of the STnc1480 gene within the *S.* Typhimurium chromosome.

#### 6.1.5 Global analysis of the post-transcriptional regulation of sRNAs

Finally, in addition to identifying the transcriptional regulatory inputs to 170 *S*. Typhimurium sRNAs, we also investigated the steady-state levels of sRNAs in mutants lacking the RNA chaperone, Hfq, and the exoribonuclease, PNPase. Hfq and PNPase are important post-transcriptional regulators of sRNA expression through controlling the stability of sRNA transcripts. We identified a core group of 46 *S*. Typhimurium sRNAs that are differentially expressed in a mutant strain lacking Hfq and are also strongly associated with Hfq, as determined by co-immunoprecipitation with Hfq (Chao et al., 2012) (Figure 3.13 and appendix III). Many of the sRNAs within this core group are likely to function via a canonical *trans*-acting mechanism of gene regulation, requiring Hfq for activity and stability. Uncharacterised *S*. Typhimurium sRNA genes within this core group, listed in appendix III, are interesting candidates for further investigation as these sRNAs appear to be functionally important under invasion-inducing conditions.

The absence of PNPase resulted in an equal number of sRNAs being up-regulated and down-regulated (Figure 3.14 A). To our knowledge there is no available PNPase co-immunoprecipitation data, thus a number of the differentially expressed sRNAs may be indirectly affected in the absence of PNPase. However, our data suggest that PNPase may play a protective role for Hfq-bound sRNAs, as has been described in *E. coli* (De Lay & Gottesman, 2011). This protective role of PNPase is limited to sRNA transcripts, as 96% of differentially expressed mRNAs were up-regulated in the absence of PNPase (Figure 3.14 B). Further experiments using Rifampicin, to distinguish between newly transcribed sRNAs and those being degraded in the presence and absence of PNPase, followed by RNA-seq are required to complement our RNA-seq based investigation of the steady-state levels of sRNAs in a Δ*pnp* mutant.

## **6.2** Future prospects

There are many possibilities for the development of this work in the future. The long term aim for this study is that the data presented here will contribute to the complete description of molecular biological networks within *S.* Typhimurium. The decreasing costs and increasing accessibility to deep sequencing technologies will result in the generation of more large-scale datasets, pertaining to cellular regulatory networks, which

will complement the network produced in this study. To the best of our knowledge, the work presented here is the first to elucidate the complex and intimate relationship between transcriptional and post-transcriptional regulators on a global scale, however, the ease at which sRNAs may now be identified will result in sRNAs becoming increasingly incorporated into established regulatory networks.

The use of regulatory mutants to investigate bacterial regulons is common practice, however, changes in gene expression may be direct or indirect. Demonstration of direct protein-DNA interactions is required to confirm direct gene regulation by a TF. Likewise, binding of a TF to a specific DNA sequence does not necessarily result in control of gene expression of nearby genes (Stormo & Zhao, 2010), thus mutagenesis-based studies, such as this, and global studies which identify protein binding sites should complement each other to determine direct pathways within regulatory networks. The move, in recent years, towards ChIP-seq technology provides the ideal technology to study proteins binding to regions corresponding to sRNA gene promoters.

Our preliminary investigations of the post-transcriptional regulation of sRNA expression, involving mutant strains that lack Hfq and PNPase, require more detailed experimentation. In addition to the, previously discussed, global analysis of sRNA stability, the effects of additional proteins which post-transcriptionally regulate the expression or activity of sRNAs, such as other components of the RNA degradosome, is required to give a more comprehensive view of post-transcriptional regulatory patterns. Furthermore, recent work to identify novel RNA binding partners is revealing interesting candidates, such as ProQ, that may play a role similar to that of Hfq in stabilising and controlling the activity of sRNAs (Jörg Vogel, Pers. Comm.). Novel mechanisms of action of the well-characterised RNA binding protein, CsrA, have recently been elucidated and suggest that, in addition to control of CsrB and CsrC (Fortune et al., 2006), CsrA may also be an important post-transcriptional regulator of the expression or activity of other sRNAs (Figueroa-Bossi et al., 2014, Jorgensen et al., 2013, Sterzenbach et al., 2013). It remains a possibility that a number of the sRNAs which were not differentially expressed in the absence of Hfq (this study), or were not found to be associated with Hfq by co-immunoprecipitation (Sittka et al., 2008, Chao et al., 2012), could function as trans-acting regulators via interaction with these, or other, alternative RNA binding partners. Co-immunoprecipitation and mutagenesis-based studies using alternative RNA binding partners will be required in the coming years.

As previously discussed, single-cell analysis is an important goal for bacterial transcriptomic-based studies because bacterial gene regulation does not occur homogenously within infected mammalian cells (Hebrard et al., 2011). Bistability of gene expression is important for virulence and bacterial survival (Sturm et al., 2011), while genetically identical but phenotypically different cells may play important roles in disease outcome and antibiotic tolerance (Helaine et al., 2014, Diard et al., 2013). Chromosomal GFP-based transcriptional fusions, coupled with flow cytometry, can accurately report the expression of a single gene on a single-cell level (Hautefort et al., 2003), but the development of global techniques, such as single-cell RNA-seq, to analyse the expression of all genes in a single bacterial cell will provide information on more subtle regulatory interactions which may be masked on a population level (Saliba et al., 2014).

Finally, further analysis of the sRNAs STnc520 and STnc1480 is required to elucidate the roles played by these sRNAs in *S.* Typhimurium pathogenesis. This study has uncovered the mechanisms of regulation of these two sRNAs by virulence-associated TFs. This study has also provided preliminary evidence about the genes targeted by these sRNAs. However, validation of a direct interaction between STnc520 and one or more of the putative target genes is required to validate our model of the role of STnc520 in protecting *S.* Typhimurium from toxic intermediate compounds during the intracellular lifestyle. A direct interaction between STnc1480 and the primary SPI1 regulator, HilA, has been validated but our hypothesis that this interaction occurs within the host epithelial cell remains to be confirmed. The putative interaction between STnc1480 and the flagellin subunit, FliC, also remains to be further investigated. Furthermore, it will be interesting to investigate if STnc1480 also regulates additional target genes that are important during intra-macrophage survival and systemic infection.

# 6.3 Concluding remarks

Overall, we have used RNA-seq to provide detailed and expanded regulons of 18 important regulatory proteins in *S.* Typhimurium. Upon publication, our transcriptomic dataset will provide a valuable community resource for researchers to investigate the regulatory pattern of their genes of interest across a panel of 18 regulatory mutants,

relative to other groups of genes which may be co-regulated. Importantly, this study also connects *S*. Typhimurium transcriptional and post-transcriptional regulatory networks by identifying the protein-factors involved in the control of sRNA expression. sRNAs may provide missing links between TFs and their target genes, while the rapidity of sRNA-mediated gene regulation extends the flexibility and dynamic range of regulatory proteins (Beisel & Storz, 2010, Mandin & Guillier, 2013, Gopel & Gorke, 2012). Analysis of our mutant RNA-seq-based transcriptomic data in conjunction with other datasets, regarding sRNA gene expression or sRNA targets, will provide a more complete picture of mixed regulatory interactions within the cell. Our particular focus on the regulons of virulence-associated regulators of *S*. Typhimurium has identified 2 sRNAs, that are apparently important for *S*. Typhimurium virulence, and may also lead to the identification of more sRNAs that contribute to the pathogenesis of *S*. Tyhimurium, taking us a step closer to the goal of elucidating all regulatory interactions and inputs involved in establishment of *S*. Typhimurium infection.

# **Bibliography**

- Abrahams, G. L. & M. Hensel, (2006) Manipulating cellular transport and immune responses: dynamic interactions between intracellular *Salmonella enterica* and its host cells. *Cell Microbiol* 8: 728-737.
- Adams, P., R. Fowler, N. Kinsella, G. Howell, M. Farris, P. Coote & C. D. O'Connor, (2001) Proteomic detection of PhoPQ- and acid-mediated repression of *Salmonella* motility. *Proteomics* 1: 597-607.
- Ahmer, B. M., J. van Reeuwijk, P. R. Watson, T. S. Wallis & F. Heffron, (1999) *Salmonella* SirA is a global regulator of genes mediating enteropathogenesis. *Mol Microbiol* **31**: 971-982.
- Aikawa, C., F. Maruyama & I. Nakagawa, (2010) The dawning era of comprehensive transcriptome analysis in cellular microbiology. *Front Microbiol* 1: 118.
- Alba, B. M. & C. A. Gross, (2004) Regulation of the *Escherichia coli* sigma-dependent envelope stress response. *Mol Microbiol* **52**: 613-619.
- Ali, S. S., J. C. Whitney, J. Stevenson, H. Robinson, P. L. Howell & W. W. Navarre, (2013) Structural insights into the regulation of foreign genes in *Salmonella* by the Hha/H-NS complex. *J Biol Chem* **288**: 13356-13369.
- Alix, E. & A. B. Blanc-Potard, (2008) Peptide-assisted degradation of the *Salmonella* MgtC virulence factor. *EMBO J* 27: 546-557.
- Altier, C., M. Suyemoto & S. D. Lawhon, (2000a) Regulation of *Salmonella enterica* serovar Typhimurium invasion genes by CsrA. *Infect Immun* **68**: 6790-6797.
- Altier, C., M. Suyemoto, A. I. Ruiz, K. D. Burnham & R. Maurer, (2000b) Characterization of two novel regulatory genes affecting *Salmonella* invasion gene expression. *Mol Microbiol* **35**: 635-646.
- Altuvia, S., (2007) Identification of bacterial small non-coding RNAs: experimental approaches. *Curr Opin Microbiol* **10**: 257-261.
- Altuvia, S. & E. G. Wagner, (2000) Switching on and off with RNA. *Proc Natl Acad Sci U S A* **97**: 9824-9826.
- Altuvia, S., D. Weinstein-Fischer, A. Zhang, L. Postow & G. Storz, (1997) A small, stable RNA induced by oxidative stress: role as a pleiotropic regulator and antimutator. *Cell* **90**: 43-53.
- Altuvia, S., A. Zhang, L. Argaman, A. Tiwari & G. Storz, (1998) The *Escherichia coli* OxyS regulatory RNA represses *fhlA* translation by blocking ribosome binding. *EMBO J* 17: 6069-6075.
- Andrade, J. M. & C. M. Arraiano, (2008) PNPase is a key player in the regulation of small RNAs that control the expression of outer membrane proteins. *RNA* **14**: 543-551.
- Andrade, J. M., V. Pobre, A. M. Matos & C. M. Arraiano, (2012) The crucial role of PNPase in the degradation of small RNAs that are not associated with Hfq. *RNA* **18**: 844-855.
- Arber, W., (2014) Horizontal Gene Transfer among Bacteria and Its Role in Biological Evolution. *Life (Basel)* **4**: 217-224.
- Aznar, S., S. Paytubi & A. Juarez, (2013) The Hha protein facilitates incorporation of horizontally acquired DNA in enteric bacteria. *Microbiology* **159**: 545-554.
- Bader, M. W., S. Sanowar, M. E. Daley, A. R. Schneider, U. Cho, W. Xu, R. E. Klevit, H. Le Moual & S. I. Miller, (2005) Recognition of antimicrobial peptides by a bacterial sensor kinase. *Cell* 122: 461-472.
- Baichoo, N. & J. D. Helmann, (2002) Recognition of DNA by Fur: a reinterpretation of the Fur box consensus sequence. *J Bacteriol* **184**: 5826-5832.
- Bajaj, V., R. L. Lucas, C. Hwang & C. A. Lee, (1996) Co-ordinate regulation of *Salmonella* Typhimurium invasion genes by environmental and regulatory factors is mediated by control of *hilA* expression. *Mol Microbiol* **22**: 703-714.
- Balbontin, R., G. Rowley, M. G. Pucciarelli, J. Lopez-Garrido, Y. Wormstone, S. Lucchini, F. Garcia-Del Portillo, J. C. Hinton & J. Casadesus, (2006) DNA adenine methylation regulates virulence gene expression in *Salmonella enterica* serovar Typhimurium. *J Bacteriol* 188: 8160-8168.

- Bang, I. S., J. P. Audia, Y. K. Park & J. W. Foster, (2002) Autoinduction of the *ompR* response regulator by acid shock and control of the *Salmonella enterica* acid tolerance response. *Mol Microbiol* 44: 1235-1250.
- Bang, I. S., J. G. Frye, M. McClelland, J. Velayudhan & F. C. Fang, (2005) Alternative sigma factor interactions in *Salmonella*: sigma and sigma promote antioxidant defences by enhancing sigma levels. *Mol Microbiol* **56**: 811-823.
- Banos, R. C., A. Vivero, S. Aznar, J. Garcia, M. Pons, C. Madrid & A. Juarez, (2009) Differential regulation of horizontally acquired and core genome genes by the bacterial modulator H-NS. *PLoS Genet* **5**: e1000513.
- Baxter, M. A., T. F. Fahlen, R. L. Wilson & B. D. Jones, (2003) HilE interacts with HilD and negatively regulates *hilA* transcription and expression of the *Salmonella enterica* serovar Typhimurium invasive phenotype. *Infect Immun* 71: 1295-1305.
- Becker, G. & R. Hengge-Aronis, (2001) What makes an *Escherichia coli* promoter sigma(S) dependent? Role of the -13/-14 nucleotide promoter positions and region 2.5 of sigma(S). *Mol Microbiol* **39**: 1153-1165.
- Behlau, I. & S. I. Miller, (1993) A PhoP-repressed gene promotes *Salmonella* Typhimurium invasion of epithelial cells. *J Bacteriol* **175**: 4475-4484.
- Beisel, C. L. & G. Storz, (2010) Base pairing small RNAs and their roles in global regulatory networks. *FEMS Microbiol Rev* **34**: 866-882.
- Beuzon, C. R., G. Banks, J. Deiwick, M. Hensel & D. W. Holden, (1999) pH-dependent secretion of SseB, a product of the SPI-2 type III secretion system of *Salmonella* Typhimurium. *Mol Microbiol* 33: 806-816.
- Bijlsma, J. J. & E. A. Groisman, (2005) The PhoP/PhoQ system controls the intramacrophage type three secretion system of *Salmonella enterica*. *Mol Microbiol* **57**: 85-96.
- Birkmann, A., R. G. Sawers & A. Bock, (1987) Involvement of the *ntrA* gene product in the anaerobic metabolism of *Escherichia coli*. *Mol Gen Genet* **210**: 535-542.
- Bittner, M., S. Saldias, C. Estevez, M. Zaldivar, C. L. Marolda, M. A. Valvano & I. Contreras, (2002) O-antigen expression in *Salmonella enterica* serovar Typhi is regulated by nitrogen availability through RpoN-mediated transcriptional control of the *rfaH* gene. *Microbiology* **148**: 3789-3799.
- Boddicker, J. D., B. M. Knosp & B. D. Jones, (2003) Transcription of the *Salmonella* invasion gene activator, *hilA*, requires HilD activation in the absence of negative regulators. *J Bacteriol* **185**: 525-533.
- Bogomolnaya, L. M., L. Aldrich, Y. Ragoza, M. Talamantes, K. D. Andrews, M. McClelland & H. L. Andrews-Polymenis, (2014) Identification of Novel Factors Involved in Modulating Motility of Salmonella enterica Serotype Typhimurium. *PLoS One* 9: e111513.
- Bonifield, H. R. & K. T. Hughes, (2003) Flagellar phase variation in *Salmonella enterica* is mediated by a posttranscriptional control mechanism. *J Bacteriol* **185**: 3567-3574.
- Bouffartigues, E., M. Buckle, C. Badaut, A. Travers & S. Rimsky, (2007) H-NS cooperative binding to high-affinity sites in a regulatory element results in transcriptional silencing. *Nat Struct Mol Biol* **14**: 441-448.
- Boysen, A., J. Moller-Jensen, B. Kallipolitis, P. Valentin-Hansen & M. Overgaard, (2010) Translational regulation of gene expression by an anaerobically induced small non-coding RNA in *Escherichia coli. J Biol Chem* **285**: 10690-10702.
- Browning, D. F. & S. J. Busby, (2004) The regulation of bacterial transcription initiation. *Nat Rev Microbiol* **2**: 57-65.
- Brumell, J. H., D. L. Goosney & B. B. Finlay, (2002) SifA, a type III secreted effector of *Salmonella* Typhimurium, directs *Salmonella*-induced filament (Sif) formation along microtubules. *Traffic* 3: 407-415.
- Buchmeier, N., S. Bossie, C. Y. Chen, F. C. Fang, D. G. Guiney & S. J. Libby, (1997) SlyA, a transcriptional regulator of *Salmonella* Typhimurium, is required for resistance to oxidative stress and is expressed in the intracellular environment of macrophages. *Infect Immun* 65: 3725-3730.

- Busch, A., A. S. Richter & R. Backofen, (2008) IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics* **24**: 2849-2856.
- Bustamante, V. H., L. C. Martinez, F. J. Santana, L. A. Knodler, O. Steele-Mortimer & J. L. Puente, (2008) HilD-mediated transcriptional cross-talk between SPI-1 and SPI-2. *Proc Natl Acad Sci U S A* **105**: 14591-14596.
- Cai, Y. H. & H. Huang, (2012) Advances in the study of protein-DNA interaction. *Amino Acids* 43: 1141-1146.
- Camacho, E. M., A. Serna, C. Madrid, S. Marques, R. Fernandez, F. de la Cruz, A. Juarez & J. Casadesus, (2005) Regulation of *finP* transcription by DNA adenine methylation in the virulence plasmid of *Salmonella enterica*. *J Bacteriol* **187**: 5691-5699.
- Cameron, A. D. & C. J. Dorman, (2012) A fundamental regulatory mechanism operating through OmpR and DNA topology controls expression of *Salmonella* pathogenicity islands SPI-1 and SPI-2. *PLoS Genet* **8**: e1002615.
- Carpenter, B. M., J. M. Whitmire & D. S. Merrell, (2009) This is not your mother's repressor: the complex role of fur in pathogenesis. *Infect Immun* 77: 2590-2601.
- Carpousis, A. J., (2007) The RNA degradosome of *Escherichia coli*: an mRNA-degrading machine assembled on RNase E. *Annu Rev Microbiol* **61**: 71-87.
- Carzaniga, T., F. Briani, S. Zangrossi, G. Merlino, P. Marchi & G. Deho, (2009) Autogenous regulation of *Escherichia coli* polynucleotide phosphorylase expression revisited. *J Bacteriol* **191**: 1738-1748.
- Casjens, S. & M. Hayden, (1988) Analysis in vivo of the bacteriophage P22 headful nuclease. *J Mol Biol* 199: 467-474.
- Chao, Y., K. Papenfort, R. Reinhardt, C. M. Sharma & J. Vogel, (2012) An atlas of Hfq-bound transcripts reveals 3' UTRs as a genomic reservoir of regulatory small RNAs. *EMBO J* 31: 4005-4019.
- Chaudhuri, R. R., E. Morgan, S. E. Peters, S. J. Pleasance, D. L. Hudson, H. M. Davies, J. Wang, P. M. van Diemen, A. M. Buckley, A. J. Bowen, G. D. Pullinger, D. J. Turner, G. C. Langridge, A. K. Turner, J. Parkhill, I. G. Charles, D. J. Maskell & M. P. Stevens, (2013) Comprehensive assignment of roles for *Salmonella* Typhimurium genes in intestinal colonization of food-producing animals. *PLoS Genet* 9: e1003456.
- Chekabab, S. M., J. Harel & C. M. Dozois, (2014) Interplay between genetic regulation of phosphate homeostasis and bacterial virulence. *Virulence* 5.
- Chen, C. C., M. Y. Chou, C. H. Huang, A. Majumder & H. Y. Wu, (2005) A cis-spreading nucleoprotein filament is responsible for the gene silencing activity found in the promoter relay mechanism. *J Biol Chem* **280**: 5101-5112.
- Chen, C. C. & H. Y. Wu, (2005) LeuO protein delimits the transcriptionally active and repressive domains on the bacterial chromosome. *J Biol Chem* **280**: 15111-15121.
- Chen, C. Y., N. A. Buchmeier, S. Libby, F. C. Fang, M. Krause & D. G. Guiney, (1995) Central regulatory role for the RpoS sigma factor in expression of *Salmonella Dublin plasmid virulence genes*. *J Bacteriol* 177: 5303-5309.
- Cheng, Z. F. & M. P. Deutscher, (2005) An important role for RNase R in mRNA decay. *Mol Cell* 17: 313-318.
- Cherepanov, P. P. & W. Wackernagel, (1995) Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene* **158**: 9-14.
- Chilcott, G. S. & K. T. Hughes, (2000) Coupling of flagellar gene expression to flagellar assembly in *Salmonella enterica* serovar Typhimurium and *Escherichia coli*. *Microbiol Mol Biol Rev* **64**: 694-708.
- Choi, E., H. Kim, H. Lee, D. Nam, J. Choi & D. Shin, (2014) The iron-sensing Fur regulator controls expression timing and levels of *Salmonella* pathogenicity island 2 genes in the course of environmental acidification. *Infect Immun* 82: 2203-2210.
- Choi, J., D. Shin, H. Yoon, J. Kim, C. R. Lee, M. Kim, Y. J. Seok & S. Ryu, (2010) *Salmonella* pathogenicity island 2 expression negatively controlled by EIIANtr-SsrB interaction is required for *Salmonella* virulence. *Proc Natl Acad Sci U S A* **107**: 20506-20511.

- Chubiz, J. E., Y. A. Golubeva, D. Lin, L. D. Miller & J. M. Slauch, (2010) FliZ regulates expression of the *Salmonella* pathogenicity island 1 invasion locus by controlling HilD protein activity in *Salmonella enterica* serovar Typhimurium. *J Bacteriol* 192: 6261-6270
- Cirillo, D. M., R. H. Valdivia, D. M. Monack & S. Falkow, (1998) Macrophage-dependent induction of the *Salmonella* pathogenicity island 2 type III secretion system and its role in intracellular survival. *Mol Microbiol* **30**: 175-188.
- Clements, M. O., S. Eriksson, A. Thompson, S. Lucchini, J. C. Hinton, S. Normark & M. Rhen, (2002) Polynucleotide phosphorylase is a global regulator of virulence and persistency in *Salmonella enterica*. *Proc Natl Acad Sci U S A* **99**: 8784-8789.
- Collazo, C. M. & J. E. Galan, (1997) The invasion-associated type III system of *Salmonella* Typhimurium directs the translocation of Sip proteins into the host cell. *Mol Microbiol* **24**: 747-756.
- Constantinidou, C., J. L. Hobman, L. Griffiths, M. D. Patel, C. W. Penn, J. A. Cole & T. W. Overton, (2006) A reassessment of the FNR regulon and transcriptomic analysis of the effects of nitrate, nitrite, NarXL, and NarQP as *Escherichia coli* K12 adapts from aerobic to anaerobic growth. *J Biol Chem* **281**: 4802-4815.
- Conway, T., J. P. Creecy, S. M. Maddox, J. E. Grissom, T. L. Conkle, T. M. Shadid, J. Teramoto, P. San Miguel, T. Shimada, A. Ishihama, H. Mori & B. L. Wanner, (2014) Unprecedented high-resolution view of bacterial operon architecture revealed by RNA sequencing. *MBio* 5: e01442-01414.
- Coornaert, A., A. Lu, P. Mandin, M. Springer, S. Gottesman & M. Guillier, (2010) MicA sRNA links the PhoP regulon to cell envelope stress. *Mol Microbiol* **76**: 467-479.
- Corcoran, C. P., D. Podkaminski, K. Papenfort, J. H. Urban, J. C. Hinton & J. Vogel, (2012) Superfolder GFP reporters validate diverse new mRNA targets of the classic porin regulator, MicF RNA. *Mol Microbiol* **84**: 428-445.
- Corpet, F., (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* **16**: 10881-10890.
- Crooks, G. E., G. Hon, J. M. Chandonia & S. E. Brenner, (2004) WebLogo: a sequence logo generator. *Genome Res* **14**: 1188-1190.
- Croucher, N. J. & N. R. Thomson, (2010) Studying bacterial transcriptomes using RNA-seq. *Curr Opin Microbiol* **13**: 619-624.
- Crump, J. A. & E. D. Mintz, (2010) Global trends in typhoid and paratyphoid Fever. *Clin Infect Dis* **50**: 241-246.
- Daniels, J. J., I. B. Autenrieth, A. Ludwig & W. Goebel, (1996) The gene *slyA* of *Salmonella* Typhimurium is required for destruction of M cells and intracellular survival but not for invasion or colonization of the murine small intestine. *Infect Immun* **64**: 5075-5084.
- Darfeuille, F., C. Unoson, J. Vogel & E. G. Wagner, (2007) An antisense RNA inhibits translation by competing with standby ribosomes. *Mol Cell* **26**: 381-392.
- Datsenko, K. A. & B. L. Wanner, (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**: 6640-6645.
- De Lay, N. & S. Gottesman, (2011) Role of polynucleotide phosphorylase in sRNA function in *Escherichia coli. RNA* 17: 1172-1189.
- Deiwick, J., T. Nikolaus, S. Erdogan & M. Hensel, (1999) Environmental regulation of *Salmonella* pathogenicity island 2 gene expression. *Mol Microbiol* **31**: 1759-1773.
- Delihas, N. & S. Forst, (2001) MicF: an antisense RNA gene involved in response of *Escherichia coli* to global stress factors. *J Mol Biol* **313**: 1-12.
- Desai, P. T., S. Porwollik, F. Long, P. Cheng, A. Wollam, V. Bhonagiri-Palsikar, K. Hallsworth-Pepin, S. W. Clifton, G. M. Weinstock & M. McClelland, (2013) Evolutionary Genomics of *Salmonella enterica* Subspecies. *MBio* 4.
- Desai, U. J. & P. K. Pfaffle, (1995) Single-step purification of a thermostable DNA polymerase expressed in *Escherichia coli*. *Biotechniques* **19**: 780-782, 784.
- Desnoyers, G., M. P. Bouchard & E. Masse, (2013) New insights into small RNA-dependent translational regulation in prokaryotes. *Trends Genet* **29**: 92-98.

- Desnoyers, G. & E. Masse, (2012) Noncanonical repression of translation initiation through small RNA recruitment of the RNA chaperone Hfq. *Genes Dev* **26**: 726-739.
- Diard, M., V. Garcia, L. Maier, M. N. Remus-Emsermann, R. R. Regoes, M. Ackermann & W. D. Hardt, (2013) Stabilization of cooperative virulence by the expression of an avirulent phenotype. *Nature* **494**: 353-356.
- Dillon, S. C., A. D. Cameron, K. Hokamp, S. Lucchini, J. C. Hinton & C. J. Dorman, (2010) Genome-wide analysis of the H-NS and Sfh regulatory networks in *Salmonella* Typhimurium identifies a plasmid-encoded transcription silencing mechanism. *Mol Microbiol* 76: 1250-1265.
- Dillon, S. C., E. Espinosa, K. Hokamp, D. W. Ussery, J. Casadesus & C. J. Dorman, (2012) LeuO is a global regulator of gene expression in *Salmonella enterica* serovar Typhimurium. *Mol Microbiol* 85: 1072-1089.
- Dong, T. & H. E. Schellhorn, (2010) Role of RpoS in virulence of pathogens. *Infect Immun* **78**: 887-897.
- Dong, T., R. Yu & H. Schellhorn, (2011) Antagonistic regulation of motility and transcriptome expression by RpoN and RpoS in *Escherichia coli*. *Mol Microbiol* **79**: 375-386.
- Dong, T. G. & J. J. Mekalanos, (2012) Characterization of the RpoN regulon reveals differential regulation of T6SS and new flagellar operons in *Vibrio cholerae* O37 strain V52. *Nucleic Acids Res* **40**: 7766-7775.
- Dorman, C. J., (2004) H-NS: a universal regulator for a dynamic genome. *Nat Rev Microbiol* **2**: 391-400.
- Dorman, C. J., S. Chatfield, C. F. Higgins, C. Hayward & G. Dougan, (1989) Characterization of porin and *ompR* mutants of a virulent strain of *Salmonella* Typhimurium: *ompR* mutants are attenuated in vivo. *Infect Immun* 57: 2136-2140.
- Dougan, G. & S. Baker, (2014) *Salmonella enterica* Serovar Typhi and the Pathogenesis of Typhoid Fever. *Annu Rev Microbiol* **68**: 317-336.
- Dower, W. J., J. F. Miller & C. W. Ragsdale, (1988) High efficiency transformation of E. coli by high voltage electroporation. *Nucleic Acids Res* **16**: 6127-6145.
- Durand, S. & G. Storz, (2010) Reprogramming of anaerobic metabolism by the FnrS small RNA. *Mol Microbiol* **75**: 1215-1231.
- Eichelberg, K., W. D. Hardt & J. E. Galan, (1999) Characterization of SprA, an AraC-like transcriptional regulator encoded within the *Salmonella* Typhimurium pathogenicity island 1. *Mol Microbiol* 33: 139-152.
- Ellermeier, C. D., J. R. Ellermeier & J. M. Slauch, (2005) HilD, HilC and RtsA constitute a feed forward loop that controls expression of the SPI1 type three secretion system regulator *hilA* in *Salmonella enterica* serovar Typhimurium. *Mol Microbiol* **57**: 691-705.
- Ellermeier, C. D. & J. M. Slauch, (2003) RtsA and RtsB coordinately regulate expression of the invasion and flagellar genes in *Salmonella enterica* serovar Typhimurium. *J Bacteriol* **185**: 5096-5108.
- Ellermeier, C. D. & J. M. Slauch, (2004) RtsA coordinately regulates DsbA and the *Salmonella* pathogenicity island 1 type III secretion system. *J Bacteriol* **186**: 68-79.
- Ellermeier, J. R. & J. M. Slauch, (2007) Adaptation to the host environment: regulation of the SPI1 type III secretion system in *Salmonella enterica* serovar Typhimurium. *Curr Opin Microbiol* **10**: 24-29.
- Ellermeier, J. R. & J. M. Slauch, (2008) Fur regulates expression of the *Salmonella* pathogenicity island 1 type III secretion system through HilD. *J Bacteriol* **190**: 476-486.
- Eriksson, S., S. Lucchini, A. Thompson, M. Rhen & J. C. Hinton, (2003) Unravelling the biology of macrophage infection by gene expression profiling of intracellular *Salmonella enterica*. *Mol Microbiol* 47: 103-118.
- Ernst, R. K., T. Guina & S. I. Miller, (2001) *Salmonella* Typhimurium outer membrane remodeling: role in resistance to host innate immunity. *Microbes Infect* **3**: 1327-1334.
- Escolar, L., J. Perez-Martin & V. de Lorenzo, (1998) Coordinated repression in vitro of the divergent *fepA-fes* promoters of *Escherichia coli* by the iron uptake regulation (Fur) protein. *J Bacteriol* **180**: 2579-2582.

- Escolar, L., J. Perez-Martin & V. de Lorenzo, (1999) Opening the iron box: transcriptional metalloregulation by the Fur protein. *J Bacteriol* **181**: 6223-6229.
- Fabrega, A. & J. Vila, (2013) *Salmonella enterica* serovar Typhimurium skills to succeed in the host: virulence and regulation. *Clin Microbiol Rev* **26**: 308-341.
- Fahlen, T. F., R. L. Wilson, J. D. Boddicker & B. D. Jones, (2001) Hha is a negative modulator of transcription of *hilA*, the *Salmonella enterica* serovar Typhimurium invasion gene transcriptional activator. *J Bacteriol* **183**: 6620-6629.
- Faith, J. J., B. Hayete, J. T. Thaden, I. Mogno, J. Wierzbowski, G. Cottarel, S. Kasif, J. J. Collins & T. S. Gardner, (2007) Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biol* 5: e8.
- Falconi, M., V. McGovern, C. Gualerzi, D. Hillyard & N. P. Higgins, (1991) Mutations altering chromosomal protein H-NS induce mini-Mu transposition. *New Biol* **3**: 615-625.
- Fang, F. C., S. J. Libby, N. A. Buchmeier, P. C. Loewen, J. Switala, J. Harwood & D. G. Guiney, (1992) The alternative sigma factor KatF (RpoS) regulates *Salmonella* virulence. *Proc Natl Acad Sci U S A* 89: 11978-11982.
- Fass, E. & E. A. Groisman, (2009) Control of *Salmonella* pathogenicity island-2 gene expression. *Curr Opin Microbiol* **12**: 199-204.
- Feasey, N. A., G. Dougan, R. A. Kingsley, R. S. Heyderman & M. A. Gordon, (2012) Invasive non-typhoidal *Salmonella* disease: an emerging and neglected tropical disease in Africa. *Lancet* **379**: 2489-2499.
- Fender, A., J. Elf, K. Hampel, B. Zimmermann & E. G. Wagner, (2010) RNAs actively cycle on the Sm-like protein Hfq. *Genes Dev* **24**: 2621-2626.
- Feng, X., R. Oropeza & L. J. Kenney, (2003) Dual regulation by phospho-OmpR of *ssrA/B* gene expression in *Salmonella* pathogenicity island 2. *Mol Microbiol* **48**: 1131-1143.
- Figueroa-Bossi, N., S. Lemire, D. Maloriol, R. Balbontin, J. Casadesus & L. Bossi, (2006) Loss of Hfq activates the sigmaE-dependent envelope stress response in *Salmonella enterica*. *Mol Microbiol* **62**: 838-852.
- Figueroa-Bossi, N., A. Schwartz, B. Guillemardet, F. D'Heygere, L. Bossi & M. Boudvillain, (2014) RNA remodeling by bacterial global regulator CsrA promotes Rho-dependent transcription termination. *Genes Dev* 28: 1239-1251.
- Fink, R. C., M. R. Evans, S. Porwollik, A. Vazquez-Torres, J. Jones-Carson, B. Troxell, S. J. Libby, M. McClelland & H. M. Hassan, (2007) FNR is a global regulator of virulence and anaerobic metabolism in *Salmonella enterica* serovar Typhimurium (ATCC 14028s). *J Bacteriol* 189: 2262-2273.
- Finlay, B. B., S. Ruschkowski & S. Dedhar, (1991) Cytoskeletal rearrangements accompanying *Salmonella* entry into epithelial cells. *J Cell Sci* **99** ( **Pt 2**): 283-296.
- Fortune, D. R., M. Suyemoto & C. Altier, (2006) Identification of CsrC and characterization of its role in epithelial cell invasion in *Salmonella enterica* serovar Typhimurium. *Infect Immun* **74**: 331-339.
- Foster, J. W., (1991) *Salmonella* acid shock proteins are required for the adaptive acid tolerance response. *J Bacteriol* **173**: 6896-6902.
- Foster, J. W. & H. K. Hall, (1990) Adaptive acidification tolerance response of *Salmonella* Typhimurium. *J Bacteriol* **172**: 771-778.
- Frohlich, K. S., K. Papenfort, A. A. Berger & J. Vogel, (2012) A conserved RpoS-dependent small RNA controls the synthesis of major porin OmpD. *Nucleic Acids Res* **40**: 3623-3640.
- Frohlich, K. S., K. Papenfort, A. Fekete & J. Vogel, (2013) A small RNA activates CFA synthase by isoform-specific mRNA stabilization. *EMBO J* **32**: 2963-2979.
- Frohlich, K. S. & J. Vogel, (2009) Activation of gene expression by small RNA. *Curr Opin Microbiol* **12**: 674-682.
- Gal-Mor, O., D. Elhadad, W. Deng, G. Rahav & B. B. Finlay, (2011) The *Salmonella enterica* PhoP directly activates the horizontally acquired SPI-2 gene *sseL* and is functionally different from a *S. bongori* ortholog. *PLoS One* 6: e20024.
- Galan, J. E., (2001) *Salmonella* interactions with host cells: type III secretion at work. *Annu Rev Cell Dev Biol* 17: 53-86.

- Galan, J. E. & D. Zhou, (2000) Striking a balance: modulation of the actin cytoskeleton by Salmonella. Proc Natl Acad Sci U S A 97: 8754-8761.
- Gantois, I., R. Ducatelle, F. Pasmans, F. Haesebrouck, I. Hautefort, A. Thompson, J. C. Hinton & F. Van Immerseel, (2006) Butyrate specifically down-regulates *Salmonella* pathogenicity island 1 gene expression. *Appl Environ Microbiol* **72**: 946-949.
- Garcia-Del Portillo, F., M. G. Pucciarelli & J. Casadesus, (1999) DNA adenine methylase mutants of *Salmonella* Typhimurium show defects in protein secretion, cell invasion, and M cell cytotoxicity. *Proc Natl Acad Sci U S A* **96**: 11578-11583.
- Garcia Vescovi, E., F. C. Soncini & E. A. Groisman, (1996) Mg2+ as an extracellular signal: environmental regulation of *Salmonella* virulence. *Cell* **84**: 165-174.
- Garmendia, J., C. R. Beuzon, J. Ruiz-Albert & D. W. Holden, (2003) The roles of SsrA-SsrB and OmpR-EnvZ in the regulation of genes encoding the *Salmonella* Typhimurium SPI-2 type III secretion system. *Microbiology* **149**: 2385-2396.
- Garsin, D. A., (2010) Ethanolamine utilization in bacterial pathogens: roles and regulation. *Nat Rev Microbiol* **8**: 290-295.
- Gerlach, R. G., N. Claudio, M. Rohde, D. Jackel, C. Wagner & M. Hensel, (2008) Cooperation of *Salmonella* pathogenicity islands 1 and 4 is required to breach epithelial barriers. *Cell Microbiol* 10: 2364-2376.
- Golubeva, Y. A., A. Y. Sadik, J. R. Ellermeier & J. M. Slauch, (2012) Integrating global regulatory input into the *Salmonella* pathogenicity island 1 type III secretion system. *Genetics* **190**: 79-90.
- Gong, H., G. P. Vu, Y. Bai, E. Chan, R. Wu, E. Yang, F. Liu & S. Lu, (2011) A *Salmonella* small non-coding RNA facilitates bacterial invasion and intracellular replication by modulating the expression of virulence factors. *PLoS Pathog* 7: e1002120.
- Gopel, Y. & B. Gorke, (2012) Rewiring two-component signal transduction with small RNAs. *Curr Opin Microbiol* **15**: 132-139.
- Gopel, Y., D. Luttmann, A. K. Heroven, B. Reichenbach, P. Dersch & B. Gorke, (2011) Common and divergent features in transcriptional control of the homologous small RNAs GlmY and GlmZ in *Enterobacteriaceae*. *Nucleic Acids Res* **39**: 1294-1309.
- Gottesman, S., (2005) Micros for microbes: non-coding regulatory RNAs in bacteria. *Trends Genet* **21**: 399-404.
- Green, J. & M. S. Paget, (2004) Bacterial redox sensors. Nat Rev Microbiol 2: 954-966.
- Grimont, P. A. D. & F.-X. Weill, (2007) Antigenic formulae of the *Salmonella* serovars, 9th edition. In: World Health Organisation Collaborating Centre for Reference and Research on *Salmonella*. Paris, France: Institute Pasteur, pp.
- Groisman, E. A. & C. Mouslim, (2006) Sensing by bacterial regulatory systems in host and non-host environments. *Nat Rev Microbiol* **4**: 705-709.
- Guillier, M. & S. Gottesman, (2006) Remodelling of the *Escherichia coli* outer membrane by two small regulatory RNAs. *Mol Microbiol* **59**: 231-247.
- Guillier, M. & S. Gottesman, (2008) The 5' end of two redundant sRNAs is involved in the regulation of multiple targets, including their own regulator. *Nucleic Acids Res* **36**: 6781-6794.
- Guiney, D. G. & J. Fierer, (2011) The Role of the spv Genes in *Salmonella Pathogenesis*. *Front Microbiol* **2**: 129.
- Gunn, J. S., C. M. Alpuche-Aranda, W. P. Loomis, W. J. Belden & S. I. Miller, (1995) Characterization of the *Salmonella* Typhimurium *pagC/pagD* chromosomal region. *J Bacteriol* 177: 5040-5047.
- Guo, M. S., T. B. Updegrove, E. B. Gogol, S. A. Shabalina, C. A. Gross & G. Storz, (2014) MicL, a new sigmaE-dependent sRNA, combats envelope stress by repressing synthesis of Lpp, the major outer membrane lipoprotein. *Genes Dev* 28: 1620-1634.
- Guzman, L. M., D. Belin, M. J. Carson & J. Beckwith, (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* **177**: 4121-4130.
- Haas, B. J., M. Chin, C. Nusbaum, B. W. Birren & J. Livny, (2012) How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes? *BMC Genomics* **13**: 734.

- Händler, K., (2014) High-throughput investigation of virulence gene regulation by sRNAs in *Salmonella* Typhimurium. In: School of Genetics and Microbiology. Trinity College Dublin, pp.
- Haraga, A., M. B. Ohlson & S. I. Miller, (2008) *Salmonellae* interplay with host cells. *Nat Rev Microbiol* **6**: 53-66.
- Hautefort, I., M. J. Proenca & J. C. Hinton, (2003) Single-copy green fluorescent protein gene fusions allow accurate measurement of *Salmonella* gene expression in vitro and during infection of mammalian cells. *Appl Environ Microbiol* **69**: 7480-7491.
- Hautefort, I., A. Thompson, S. Eriksson-Ygberg, M. L. Parker, S. Lucchini, V. Danino, R. J. Bongaerts, N. Ahmad, M. Rhen & J. C. Hinton, (2008) During infection of epithelial cells *Salmonella enterica* serovar Typhimurium undergoes a time-dependent transcriptional adaptation that results in simultaneous expression of three type 3 secretion systems. *Cell Microbiol* 10: 958-984.
- Hebrard, M., C. Kroger, S. K. Sivasankaran, K. Handler & J. C. Hinton, (2011) The challenge of relating gene expression to the virulence of *Salmonella enterica* serovar Typhimurium. *Curr Opin Biotechnol* **22**: 200-210.
- Hebrard, M., C. Kroger, S. Srikumar, A. Colgan, K. Handler & J. C. Hinton, (2012) sRNAs and the virulence of *Salmonella enterica* serovar Typhimurium. *RNA Biol* 9: 437-445.
- Helaine, S., A. M. Cheverton, K. G. Watson, L. M. Faure, S. A. Matthews & D. W. Holden, (2014) Internalization of *Salmonella* by macrophages induces formation of nonreplicating persisters. *Science* **343**: 204-208.
- Hengge-Aronis, R., (2002) Signal transduction and regulatory mechanisms involved in control of the sigma(S) (RpoS) subunit of RNA polymerase. *Microbiol Mol Biol Rev* **66**: 373-395, table of contents.
- Hensel, M., (2000) Salmonella pathogenicity island 2. Mol Microbiol 36: 1015-1023.
- Hinton, J. C., D. S. Santos, A. Seirafi, C. S. Hulton, G. D. Pavitt & C. F. Higgins, (1992) Expression and mutational analysis of the nucleoid-associated protein H-NS of *Salmonella* Typhimurium. *Mol Microbiol* 6: 2327-2337.
- Hobbs, E. C., J. L. Astarita & G. Storz, (2010) Small RNAs and small proteins involved in resistance to cell envelope stress and acid shock in *Escherichia coli*: analysis of a barcoded mutant collection. *J Bacteriol* **192**: 59-67.
- Hodges, A. P., D. Dai, Z. Xiang, P. Woolf, C. Xi & Y. He, (2010) Bayesian network expansion identifies new ROS and biofilm regulators. *PLoS One* **5**: e9513.
- Hodgkinson, V. & M. J. Petris, (2012) Copper homeostasis at the host-pathogen interface. *J Biol Chem* **287**: 13549-13555.
- Hoffmann, S., C. Otto, S. Kurtz, C. M. Sharma, P. Khaitovich, J. Vogel, P. F. Stadler & J. Hackermuller, (2009) Fast mapping of short sequences with mismatches, insertions and deletions using index structures. *PLoS Comput Biol* **5**: e1000502.
- Holmqvist, E., J. Reimegard, M. Sterk, N. Grantcharova, U. Romling & E. G. Wagner, (2010) Two antisense RNAs target the transcriptional regulator CsgD to inhibit curli synthesis. *EMBO J* **29**: 1840-1850.
- Holmqvist, E., C. Unoson, J. Reimegard & E. G. Wagner, (2012) A mixed double negative feedback loop between the sRNA MicF and the global regulator Lrp. *Mol Microbiol* **84**: 414-427.
- Huang, Y., M. Suyemoto, C. D. Garner, K. M. Cicconi & C. Altier, (2008) Formate acts as a diffusible signal to induce *Salmonella* invasion. *J Bacteriol* **190**: 4233-4241.
- Humphreys, S., A. Stevenson, A. Bacon, A. B. Weinhardt & M. Roberts, (1999) The alternative sigma factor, sigmaE, is critically important for the virulence of *Salmonella* Typhimurium. *Infect Immun* 67: 1560-1568.
- Ikebe, T., S. Iyoda & K. Kutsukake, (1999) Structure and expression of the *fliA* operon of *Salmonella* Typhimurium. *Microbiology* **145** ( **Pt 6**): 1389-1396.
- Jacobsen, A., R. S. Hendriksen, F. M. Aaresturp, D. W. Ussery & C. Friis, (2011) The *Salmonella enterica* pan-genome. *Microb Ecol* **62**: 487-504.

- Jakomin, M., D. Chessa, A. J. Baumler & J. Casadesus, (2008) Regulation of the *Salmonella enterica std* fimbrial operon by DNA adenine methylation, SeqA, and HdfR. *J Bacteriol* **190**: 7406-7413.
- Jamieson, D. J. & C. F. Higgins, (1986) Two genetically distinct pathways for transcriptional regulation of anaerobic gene expression in *Salmonella* Typhimurium. *J Bacteriol* **168**: 389-397.
- Johnston, C., D. A. Pegues, C. J. Hueck, A. Lee & S. I. Miller, (1996) Transcriptional activation of *Salmonella* Typhimurium invasion genes by a member of the phosphorylated response-regulator superfamily. *Mol Microbiol* **22**: 715-727.
- Jones, A. M., A. Goodwill & T. Elliott, (2006) Limited role for the DsrA and RprA regulatory RNAs in *rpoS* regulation in *Salmonella enterica*. *J Bacteriol* **188**: 5077-5088.
- Jones, B. D., (2005) *Salmonella* invasion gene regulation: a story of environmental awareness. *J Microbiol* **43 Spec No**: 110-117.
- Jones, G. W., L. A. Richardson & D. Uhlman, (1981) The invasion of HeLa cells by *Salmonella* Typhimurium: reversible and irreversible bacterial attachment and the role of bacterial motility. *J Gen Microbiol* **127**: 351-360.
- Jorgensen, M. G., J. S. Nielsen, A. Boysen, T. Franch, J. Moller-Jensen & P. Valentin-Hansen, (2012) Small regulatory RNAs control the multi-cellular adhesive lifestyle of *Escherichia coli. Mol Microbiol* **84**: 36-50.
- Jorgensen, M. G., M. K. Thomason, J. Havelund, P. Valentin-Hansen & G. Storz, (2013) Dual function of the McaS small RNA in controlling biofilm formation. *Genes Dev* 27: 1132-1145.
- Kadner, R. J., (2005) Regulation by iron: RNA rules the rust. J Bacteriol 187: 6870-6873.
- Kage, H., A. Takaya, M. Ohya & T. Yamamoto, (2008) Coordinated regulation of expression of *Salmonella* pathogenicity island 1 and flagellar type III secretion systems by ATP-dependent ClpXP protease. *J Bacteriol* **190**: 2470-2478.
- Kato, A. & E. A. Groisman, (2008) The PhoQ/PhoP regulatory network of *Salmonella enterica*. *Adv Exp Med Biol* **631**: 7-21.
- Kenney, L. J., (2002) Structure/function relationships in OmpR and other winged-helix transcription factors. *Curr Opin Microbiol* **5**: 135-141.
- Kery, M. B., M. Feldman, J. Livny & B. Tjaden, (2014) TargetRNA2: identifying targets of small regulatory RNAs in bacteria. *Nucleic Acids Res* **42**: W124-129.
- Kiley, P. J. & H. Beinert, (2003) The role of Fe-S proteins in sensing and regulation in bacteria. *Curr Opin Microbiol* **6**: 181-185.
- Kimbrough, T. G. & S. I. Miller, (2000) Contribution of *Salmonella* Typhimurium type III secretion components to needle complex formation. *Proc Natl Acad Sci U S A* **97**: 11008-11013.
- Knodler, L. A., V. Nair & O. Steele-Mortimer, (2014) Quantitative assessment of cytosolic *Salmonella* in epithelial cells. *PLoS One* **9**: e84681.
- Knodler, L. A. & O. Steele-Mortimer, (2003) Taking possession: biogenesis of the *Salmonella*-containing vacuole. *Traffic* **4**: 587-599.
- Kong, W., N. Weatherspoon & Y. Shi, (2008) Molecular mechanism for establishment of signal-dependent regulation in the PhoP/PhoQ system. *J Biol Chem* **283**: 16612-16621.
- Kowarz, L., C. Coynault, V. Robbe-Saule & F. Norel, (1994) The *Salmonella* Typhimurium *katF* (*rpoS*) gene: cloning, nucleotide sequence, and regulation of *spvR* and *spvABCD* virulence plasmid genes. *J Bacteriol* **176**: 6852-6860.
- Kredich, N. M., (1992) The molecular basis for positive regulation of *cys* promoters in *Salmonella* Typhimurium and *Escherichia coli*. *Mol Microbiol* **6**: 2747-2753.
- Krieger, V., D. Liebl, Y. Zhang, R. Rajashekar, P. Chlanda, K. Giesker, D. Chikkaballi & M. Hensel, (2014) Reorganization of the endosomal system in *Salmonella*-infected cells: the ultrastructure of *Salmonella*-induced tubular compartments. *PLoS Pathog* **10**: e1004374.
- Kröger, C., A. Colgan, S. Srikumar, K. Handler, S. K. Sivasankaran, D. L. Hammarlof, R. Canals, J. E. Grissom, T. Conway, K. Hokamp & J. C. Hinton, (2013) An infection-relevant transcriptomic compendium for *Salmonella enterica* Serovar Typhimurium. *Cell Host Microbe* 14: 683-695.

- Kröger, C., S. C. Dillon, A. D. Cameron, K. Papenfort, S. K. Sivasankaran, K. Hokamp, Y. Chao, A. Sittka, M. Hebrard, K. Handler, A. Colgan, P. Leekitcharoenphon, G. C. Langridge, A. J. Lohan, B. Loftus, S. Lucchini, D. W. Ussery, C. J. Dorman, N. R. Thomson, J. Vogel & J. C. Hinton, (2012) The transcriptional landscape and small RNAs of *Salmonella enterica* serovar Typhimurium. *Proc Natl Acad Sci U S A* 109: E1277-1286.
- Laemmli, U. K., (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-685.
- Lambden, P. R. & J. R. Guest, (1976) Mutants of *Escherichia coli* K12 unable to use fumarate as an anaerobic electron acceptor. *J Gen Microbiol* **97**: 145-160.
- Lawhon, S. D., R. Maurer, M. Suyemoto & C. Altier, (2002) Intestinal short-chain fatty acids alter *Salmonella* Typhimurium invasion gene expression and virulence through BarA/SirA. *Mol Microbiol* **46**: 1451-1464.
- Lease, R. A. & M. Belfort, (2000) A trans-acting RNA as a control switch in *Escherichia coli*: DsrA modulates function by forming alternative structures. *Proc Natl Acad Sci U S A* **97**: 9919-9924.
- Lease, R. A., M. E. Cusick & M. Belfort, (1998) Riboregulation in *Escherichia coli*: DsrA RNA acts by RNA:RNA interactions at multiple loci. *Proc Natl Acad Sci U S A* **95**: 12456-12461.
- Lee, D. J., S. D. Minchin & S. J. Busby, (2012) Activating transcription in bacteria. *Annu Rev Microbiol* **66**: 125-152.
- Lee, E. J. & E. A. Groisman, (2010) An antisense RNA that governs the expression kinetics of a multifunctional virulence gene. *Mol Microbiol* **76**: 1020-1033.
- Lee, Y. H., S. Kim, J. D. Helmann, B. H. Kim & Y. K. Park, (2013) RaoN, a small RNA encoded within *Salmonella* pathogenicity island-11, confers resistance to macrophage-induced stress. *Microbiology* **159**: 1366-1378.
- Lefebre, M. D. & J. E. Galan, (2014) The inner rod protein controls substrate switching and needle length in a *Salmonella* type III secretion system. *Proc Natl Acad Sci U S A* **111**: 817-822.
- Lejona, S., A. Aguirre, M. L. Cabeza, E. Garcia Vescovi & F. C. Soncini, (2003) Molecular characterization of the Mg2+-responsive PhoP-PhoQ regulon in *Salmonella enterica*. *J Bacteriol* **185**: 6287-6294.
- Lennox, E. S., (1955) Transduction of linked genetic characters of the host by bacteriophage P1. *Virology* 1: 190-206.
- Leonhartsberger, S., I. Korsa & A. Bock, (2002) The molecular biology of formate metabolism in enterobacteria. *J Mol Microbiol Biotechnol* **4**: 269-276.
- Levi-Meyrueis, C., V. Monteil, O. Sismeiro, M. A. Dillies, M. Monot, B. Jagla, J. Y. Coppee, B. Dupuy & F. Norel, (2014) Expanding the RpoS/sigmaS-Network by RNA Sequencing and Identification of sigmaS-Controlled Small RNAs in *Salmonella*. *PLoS One* **9**: e96918.
- Levin, J. Z., M. Yassour, X. Adiconis, C. Nusbaum, D. A. Thompson, N. Friedman, A. Gnirke & A. Regev, (2010) Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nat Methods* **7**: 709-715.
- Li, M. Z. & S. J. Elledge, (2007) Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat Methods* **4**: 251-256.
- Libby, S. J., W. Goebel, A. Ludwig, N. Buchmeier, F. Bowe, F. C. Fang, D. G. Guiney, J. G. Songer & F. Heffron, (1994) A cytolysin encoded by *Salmonella* is required for survival within macrophages. *Proc Natl Acad Sci U S A* **91**: 489-493.
- Linehan, S. A., A. Rytkonen, X. J. Yu, M. Liu & D. W. Holden, (2005) SlyA regulates function of *Salmonella* pathogenicity island 2 (SPI-2) and expression of SPI-2-associated genes. *Infect Immun* 73: 4354-4362.
- Lithgow, J. K., F. Haider, I. S. Roberts & J. Green, (2007) Alternate SlyA and H-NS nucleoprotein complexes control *hlyE* expression in *Escherichia coli* K-12. *Mol Microbiol* **66**: 685-698.
- Liu, J. M. & A. Camilli, (2011) Discovery of bacterial sRNAs by high-throughput sequencing. *Methods Mol Biol* **733**: 63-79.

- Lober, S., D. Jackel, N. Kaiser & M. Hensel, (2006) Regulation of *Salmonella* pathogenicity island 2 genes by independent environmental signals. *Int J Med Microbiol* **296**: 435-447.
- Lonetto, M. A., V. Rhodius, K. Lamberg, P. Kiley, S. Busby & C. Gross, (1998) Identification of a contact site for different transcription activators in region 4 of the *Escherichia coli* RNA polymerase sigma70 subunit. *J Mol Biol* **284**: 1353-1365.
- Lopez-Garrido, J. & J. Casadesus, (2010) Regulation of *Salmonella enterica* pathogenicity island 1 by DNA adenine methylation. *Genetics* **184**: 637-649.
- Low, D. A. & J. Casadesus, (2008) Clocks and switches: bacterial gene regulation by DNA adenine methylation. *Curr Opin Microbiol* 11: 106-112.
- Lucas, R. L. & C. A. Lee, (2001) Roles of *hilC* and *hilD* in regulation of *hilA* expression in *Salmonella enterica* serovar Typhimurium. *J Bacteriol* **183**: 2733-2745.
- Lucas, R. L., C. P. Lostroh, C. C. DiRusso, M. P. Spector, B. L. Wanner & C. A. Lee, (2000) Multiple factors independently regulate *hilA* and invasion gene expression in *Salmonella enterica* serovar Typhimurium. *J Bacteriol* **182**: 1872-1882.
- Lucchini, S., P. McDermott, A. Thompson & J. C. Hinton, (2009) The H-NS-like protein StpA represses the RpoS (sigma 38) regulon during exponential growth of *Salmonella* Typhimurium. *Mol Microbiol* **74**: 1169-1186.
- Lucchini, S., G. Rowley, M. D. Goldberg, D. Hurd, M. Harrison & J. C. Hinton, (2006) H-NS mediates the silencing of laterally acquired genes in bacteria. *PLoS Pathog* 2: e81.
- Madrid, C., C. Balsalobre, J. Garcia & A. Juarez, (2007) The novel Hha/YmoA family of nucleoid-associated proteins: use of structural mimicry to modulate the activity of the H-NS family of proteins. *Mol Microbiol* **63**: 7-14.
- Main-Hester, K. L., K. M. Colpitts, G. A. Thomas, F. C. Fang & S. J. Libby, (2008) Coordinate regulation of *Salmonella* pathogenicity island 1 (SPI1) and SPI4 in *Salmonella enterica* serovar Typhimurium. *Infect Immun* 76: 1024-1035.
- Majdalani, N., S. Chen, J. Murrow, K. St John & S. Gottesman, (2001) Regulation of RpoS by a novel small RNA: the characterization of RprA. *Mol Microbiol* **39**: 1382-1394.
- Majdalani, N., C. Cunning, D. Sledjeski, T. Elliott & S. Gottesman, (1998) DsrA RNA regulates translation of RpoS message by an anti-antisense mechanism, independent of its action as an antisilencer of transcription. *Proc Natl Acad Sci U S A* **95**: 12462-12467.
- Majdalani, N., D. Hernandez & S. Gottesman, (2002) Regulation and mode of action of the second small RNA activator of RpoS translation, RprA. *Mol Microbiol* **46**: 813-826.
- Majowicz, S. E., J. Musto, E. Scallan, F. J. Angulo, M. Kirk, S. J. O'Brien, T. F. Jones, A. Fazil & R. M. Hoekstra, (2010) The global burden of nontyphoidal *Salmonella* gastroenteritis. *Clin Infect Dis* **50**: 882-889.
- Makino, K., M. Amemura, S. K. Kim, A. Nakata & H. Shinagawa, (1993) Role of the sigma 70 subunit of RNA polymerase in transcriptional activation by activator protein PhoB in *Escherichia coli. Genes Dev* 7: 149-160.
- Mandin, P. & S. Gottesman, (2010) Integrating anaerobic/aerobic sensing and the general stress response through the ArcZ small RNA. *EMBO J* **29**: 3094-3107.
- Mandin, P. & M. Guillier, (2013) Expanding control in bacteria: interplay between small RNAs and transcriptional regulators to control gene expression. *Curr Opin Microbiol* **16**: 125-132.
- Marinus, M. G. & J. Casadesus, (2009) Roles of DNA adenine methylation in host-pathogen interactions: mismatch repair, transcriptional regulation, and more. *FEMS Microbiol Rev* **33**: 488-503.
- Marteyn, B., N. P. West, D. F. Browning, J. A. Cole, J. G. Shaw, F. Palm, J. Mounier, M. C. Prevost, P. Sansonetti & C. M. Tang, (2010) Modulation of *Shigella* virulence in response to available oxygen in vivo. *Nature* **465**: 355-358.
- Martinez, L. C., M. M. Banda, M. Fernandez-Mora, F. J. Santana & V. H. Bustamante, (2014) HilD Induces Expression of SPI-2 Genes by Displacing the Global Negative Regulator H-NS from *ssrAB*. *J Bacteriol*.
- Martinez, L. C., H. Yakhnin, M. I. Camacho, D. Georgellis, P. Babitzke, J. L. Puente & V. H. Bustamante, (2011) Integration of a complex regulatory cascade involving the SirA/BarA

- and Csr global regulatory systems that controls expression of the *Salmonella* SPI-1 and SPI-2 virulence regulons through HilD. *Mol Microbiol* **80**: 1637-1656.
- Masse, E. & S. Gottesman, (2002) A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proc Natl Acad Sci U S A* **99**: 4620-4625.
- Masse, E., C. K. Vanderpool & S. Gottesman, (2005) Effect of RyhB small RNA on global iron use in *Escherichia coli*. *J Bacteriol* **187**: 6962-6971.
- McClelland, M., K. E. Sanderson, S. W. Clifton, P. Latreille, S. Porwollik, A. Sabo, R. Meyer, T. Bieri, P. Ozersky, M. McLellan, C. R. Harkins, C. Wang, C. Nguyen, A. Berghoff, G. Elliott, S. Kohlberg, C. Strong, F. Du, J. Carter, C. Kremizki, D. Layman, S. Leonard, H. Sun, L. Fulton, W. Nash, T. Miner, P. Minx, K. Delehaunty, C. Fronick, V. Magrini, M. Nhan, W. Warren, L. Florea, J. Spieth & R. K. Wilson, (2004) Comparison of genome degradation in Paratyphi A and Typhi, human-restricted serovars of Salmonella enterica that cause typhoid. Nat Genet 36: 1268-1274.
- McClelland, M., K. E. Sanderson, J. Spieth, S. W. Clifton, P. Latreille, L. Courtney, S. Porwollik, J. Ali, M. Dante, F. Du, S. Hou, D. Layman, S. Leonard, C. Nguyen, K. Scott, A. Holmes, N. Grewal, E. Mulvaney, E. Ryan, H. Sun, L. Florea, W. Miller, T. Stoneking, M. Nhan, R. Waterston & R. K. Wilson, (2001) Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature* 413: 852-856.
- McCullen, C. A., J. N. Benhammou, N. Majdalani & S. Gottesman, (2010) Mechanism of positive regulation by DsrA and RprA small noncoding RNAs: pairing increases translation and protects *rpoS* mRNA from degradation. *J Bacteriol* **192**: 5559-5571.
- McDermott, J. E., R. C. Taylor, H. Yoon & F. Heffron, (2009) Bottlenecks and hubs in inferred networks are important for virulence in *Salmonella* Typhimurium. *J Comput Biol* **16**: 169-180.
- Merrick, M. J., (1993) In a class of its own--the RNA polymerase sigma factor sigma 54 (sigma N). *Mol Microbiol* **10**: 903-909.
- Miller, J. H., (1972) A Short Course in Bacterial Genetics: A Laboratory Manual and Handbook for *Escherichia coli* and Related Bacteria. *Cold Spring Harbour Press*.
- Miticka, H., G. Rowley, B. Rezuchova, D. Homerova, S. Humphreys, J. Farn, M. Roberts & J. Kormanec, (2003) Transcriptional analysis of the *rpoE* gene encoding extracytoplasmic stress response sigma factor sigmaE in *Salmonella enterica* serovar Typhimurium. *FEMS Microbiol Lett* **226**: 307-314.
- Mizuno, T., M. Y. Chou & M. Inouye, (1984) A unique mechanism regulating gene expression: translational inhibition by a complementary RNA transcript (micRNA). *Proc Natl Acad Sci U S A* **81**: 1966-1970.
- Mohanty, B. K., V. F. Maples & S. R. Kushner, (2004) The Sm-like protein Hfq regulates polyadenylation dependent mRNA decay in *Escherichia coli*. *Mol Microbiol* **54**: 905-920.
- Moll, I., T. Afonyushkin, O. Vytvytska, V. R. Kaberdin & U. Blasi, (2003) Coincident Hfq binding and RNase E cleavage sites on mRNA and small regulatory RNAs. RNA 9: 1308-1314.
- Monteiro, C., K. Papenfort, K. Hentrich, I. Ahmad, S. Le Guyon, R. Reimann, N. Grantcharova & U. Romling, (2012) Hfq and Hfq-dependent small RNAs are major contributors to multicellular development in *Salmonella enterica* serovar Typhimurium. *RNA Biol* 9: 489-502.
- Moon, K. & S. Gottesman, (2009) A PhoQ/P-regulated small RNA regulates sensitivity of *Escherichia coli* to antimicrobial peptides. *Mol Microbiol* **74**: 1314-1330.
- Moon, K. & S. Gottesman, (2011) Competition among Hfq-binding small RNAs in *Escherichia coli*. *Mol Microbiol* **82**: 1545-1562.
- Morett, E. & L. Segovia, (1993) The sigma 54 bacterial enhancer-binding protein family: mechanism of action and phylogenetic relationship of their functional domains. *J Bacteriol* **175**: 6067-6074.
- Mortazavi, A., B. A. Williams, K. McCue, L. Schaeffer & B. Wold, (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5**: 621-628.
- Murphy, K. C., (1998) Use of bacteriophage lambda recombination functions to promote gene replacement in *Escherichia coli*. *J Bacteriol* **180**: 2063-2071.

- Mutz, K. O., A. Heilkenbrinker, M. Lonne, J. G. Walter & F. Stahl, (2013) Transcriptome analysis using next-generation sequencing. *Curr Opin Biotechnol* **24**: 22-30.
- Myers, K. S., H. Yan, I. M. Ong, D. Chung, K. Liang, F. Tran, S. Keles, R. Landick & P. J. Kiley, (2013) Genome-scale analysis of *Escherichia coli* FNR reveals complex features of transcription factor binding. *PLoS Genet* 9: e1003565.
- Navarre, W. W., T. A. Halsey, D. Walthers, J. Frye, M. McClelland, J. L. Potter, L. J. Kenney, J. S. Gunn, F. C. Fang & S. J. Libby, (2005) Co-regulation of *Salmonella enterica* genes required for virulence and resistance to antimicrobial peptides by SlyA and PhoP/PhoQ. *Mol Microbiol* 56: 492-508.
- Navarre, W. W., M. McClelland, S. J. Libby & F. C. Fang, (2007) Silencing of xenogeneic DNA by H-NS-facilitation of lateral gene transfer in bacteria by a defense system that recognizes foreign DNA. *Genes Dev* **21**: 1456-1471.
- Navarre, W. W., S. Porwollik, Y. Wang, M. McClelland, H. Rosen, S. J. Libby & F. C. Fang, (2006) Selective silencing of foreign DNA with low GC content by the H-NS protein in *Salmonella*. *Science* **313**: 236-238.
- Nicol, J. W., G. A. Helt, S. G. Blanchard, Jr., A. Raja & A. E. Loraine, (2009) The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. *Bioinformatics* **25**: 2730-2731.
- Norte, V. A., M. R. Stapleton & J. Green, (2003) PhoP-responsive expression of the *Salmonella enterica* serovar Typhimurium *slyA* gene. *J Bacteriol* **185**: 3508-3514.
- Ochman, H. & E. A. Groisman, (1996) Distribution of pathogenicity islands in *Salmonella* spp. *Infect Immun* **64**: 5410-5412.
- Okada, N., Y. Oi, M. Takeda-Shitaka, K. Kanou, H. Umeyama, T. Haneda, T. Miki, S. Hosoya & H. Danbara, (2007) Identification of amino acid residues of *Salmonella* SlyA that are critical for transcriptional regulation. *Microbiology* **153**: 548-560.
- Olekhnovich, I. N. & R. J. Kadner, (2002) DNA-binding activities of the HilC and HilD virulence regulatory proteins of *Salmonella enterica* serovar Typhimurium. *J Bacteriol* **184**: 4148-4160.
- Olekhnovich, I. N. & R. J. Kadner, (2004) Contribution of the RpoA C-terminal domain to stimulation of the *Salmonella enterica hilA* promoter by HilC and HilD. *J Bacteriol* **186**: 3249-3253.
- Olekhnovich, I. N. & R. J. Kadner, (2006) Crucial roles of both flanking sequences in silencing of the *hilA* promoter in *Salmonella enterica*. *J Mol Biol* **357**: 373-386.
- Olekhnovich, I. N. & R. J. Kadner, (2007) Role of nucleoid-associated proteins Hha and H-NS in expression of *Salmonella enterica* activators HilD, HilC, and RtsA required for cell invasion. *J Bacteriol* **189**: 6882-6890.
- Oren, Y., M. B. Smith, N. I. Johns, M. Kaplan Zeevi, D. Biran, E. Z. Ron, J. Corander, H. H. Wang, E. J. Alm & T. Pupko, (2014) Transfer of noncoding DNA drives regulatory rewiring in bacteria. *Proc Natl Acad Sci U S A*.
- Osborne, S. E. & B. K. Coombes, (2011) Transcriptional priming of *Salmonella Pathogenicity* Island-2 precedes cellular invasion. *PLoS One* **6**: e21648.
- Osborne, S. E., D. Walthers, A. M. Tomljenovic, D. T. Mulder, U. Silphaduang, N. Duong, M. J. Lowden, M. E. Wickham, R. F. Waller, L. J. Kenney & B. K. Coombes, (2009) Pathogenic adaptation of intracellular bacteria by rewiring a cis-regulatory input function. *Proc Natl Acad Sci U S A* **106**: 3982-3987.
- Osman, D., C. J. Patterson, K. Bailey, K. Fisher, N. J. Robinson, S. E. Rigby & J. S. Cavet, (2013) The copper supply pathway to a *Salmonella* Cu,Zn-superoxide dismutase (SodCII) involves P(1B)-type ATPase copper efflux and periplasmic CueP. *Mol Microbiol* 87: 466-477.
- Osman, D., K. J. Waldron, H. Denton, C. M. Taylor, A. J. Grant, P. Mastroeni, N. J. Robinson & J. S. Cavet, (2010) Copper homeostasis in *Salmonella* is atypical and copper-CueP is a major periplasmic metal complex. *J Biol Chem* **285**: 25259-25268.
- Ozsolak, F. & P. M. Milos, (2011) RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet* **12**: 87-98.

- Ozsolak, F., A. R. Platt, D. R. Jones, J. G. Reifenberger, L. E. Sass, P. McInerney, J. F. Thompson, J. Bowers, M. Jarosz & P. M. Milos, (2009) Direct RNA sequencing. *Nature* 461: 814-818.
- Padalon-Brauch, G., R. Hershberg, M. Elgrably-Weiss, K. Baruch, I. Rosenshine, H. Margalit & S. Altuvia, (2008) Small RNAs encoded within genetic islands of *Salmonella* Typhimurium show host-induced expression and role in virulence. *Nucleic Acids Res* **36**: 1913-1927.
- Paetzold, S., S. Lourido, B. Raupach & A. Zychlinsky, (2007) *Shigella flexneri* phagosomal escape is independent of invasion. *Infect Immun* **75**: 4826-4830.
- Papenfort, K., M. Bouvier, F. Mika, C. M. Sharma & J. Vogel, (2010) Evidence for an autonomous 5' target recognition domain in an Hfq-associated small RNA. *Proc Natl Acad Sci U S A* **107**: 20435-20440.
- Papenfort, K., V. Pfeiffer, S. Lucchini, A. Sonawane, J. C. Hinton & J. Vogel, (2008) Systematic deletion of *Salmonella* small RNA genes identifies CyaR, a conserved CRP-dependent riboregulator of OmpX synthesis. *Mol Microbiol* **68**: 890-906.
- Papenfort, K., V. Pfeiffer, F. Mika, S. Lucchini, J. C. Hinton & J. Vogel, (2006) SigmaE-dependent small RNAs of *Salmonella* respond to membrane stress by accelerating global *omp* mRNA decay. *Mol Microbiol* **62**: 1674-1688.
- Papenfort, K., D. Podkaminski, J. C. Hinton & J. Vogel, (2012) The ancestral SgrS RNA discriminates horizontally acquired *Salmonella* mRNAs through a single G-U wobble pair. *Proc Natl Acad Sci U S A* **109**: E757-764.
- Papenfort, K., N. Said, T. Welsink, S. Lucchini, J. C. Hinton & J. Vogel, (2009) Specific and pleiotropic patterns of mRNA regulation by ArcZ, a conserved, Hfq-dependent small RNA. *Mol Microbiol* **74**: 139-158.
- Papenfort, K., Y. Sun, M. Miyakoshi, C. K. Vanderpool & J. Vogel, (2013) Small RNA-mediated activation of sugar phosphatase mRNA regulates glucose homeostasis. *Cell* **153**: 426-437.
- Perez, J. C., T. Latifi & E. A. Groisman, (2008) Overcoming H-NS-mediated transcriptional silencing of horizontally acquired genes by the PhoP and SlyA proteins in *Salmonella enterica*. *J Biol Chem* **283**: 10773-10783.
- Perkins, T. T., M. R. Davies, E. J. Klemm, G. Rowley, T. Wileman, K. James, T. Keane, D. Maskell, J. C. Hinton, G. Dougan & R. A. Kingsley, (2013) ChIP-seq and transcriptome analysis of the OmpR regulon of *Salmonella enterica* serovars Typhi and Typhimurium reveals accessory genes implicated in host colonization. *Mol Microbiol* 87: 526-538.
- Petrone, B. L., A. M. Stringer & J. T. Wade, (2014) Identification of HilD-regulated genes in *Salmonella enterica* serovar Typhimurium. *J Bacteriol* **196**: 1094-1101.
- Pfeiffer, V., K. Papenfort, S. Lucchini, J. C. Hinton & J. Vogel, (2009) Coding sequence targeting by MicC RNA reveals bacterial mRNA silencing downstream of translational initiation. *Nat Struct Mol Biol* **16**: 840-846.
- Pfeiffer, V., A. Sittka, R. Tomer, K. Tedin, V. Brinkmann & J. Vogel, (2007) A small non-coding RNA of the invasion gene island (SPI-1) represses outer membrane protein synthesis from the *Salmonella* core genome. *Mol Microbiol* **66**: 1174-1191.
- Pontel, L. B. & F. C. Soncini, (2009) Alternative periplasmic copper-resistance mechanisms in Gram negative bacteria. *Mol Microbiol* **73**: 212-225.
- Porwollik, S. & M. McClelland, (2003) Lateral gene transfer in *Salmonella*. *Microbes Infect* 5: 977-989.
- Prevost, K., H. Salvail, G. Desnoyers, J. F. Jacques, E. Phaneuf & E. Masse, (2007) The small RNA RyhB activates the translation of *shiA* mRNA encoding a permease of shikimate, a compound involved in siderophore synthesis. *Mol Microbiol* **64**: 1260-1273.
- Prieto, A. I., F. Ramos-Morales & J. Casadesus, (2004) Bile-induced DNA damage in *Salmonella enterica*. *Genetics* **168**: 1787-1794.
- Prost, L. R., M. E. Daley, V. Le Sage, M. W. Bader, H. Le Moual, R. E. Klevit & S. I. Miller, (2007) Activation of the bacterial sensor kinase PhoQ by acidic pH. *Mol Cell* **26**: 165-174.
- Quackenbush, J., (2002) Microarray data normalization and transformation. *Nat Genet* **32 Suppl**: 496-501.

- Quackenbush, J., (2003) Genomics. Microarrays--guilt by association. Science 302: 240-241.
- Quan, J. A., B. L. Schneider, I. T. Paulsen, M. Yamada, N. M. Kredich & M. H. Saier, Jr., (2002) Regulation of carbon utilization by sulfur availability in *Escherichia coli* and *Salmonella* Typhimurium. *Microbiology* **148**: 123-131.
- Quinn, H. J., A. D. Cameron & C. J. Dorman, (2014) Bacterial regulon evolution: distinct responses and roles for the identical OmpR proteins of *Salmonella* Typhimurium and *Escherichia coli* in the acid stress response. *PLoS Genet* **10**: e1004215.
- Raffatellu, M., R. P. Wilson, D. Chessa, H. Andrews-Polymenis, Q. T. Tran, S. Lawhon, S. Khare, L. G. Adams & A. J. Baumler, (2005) SipA, SopA, SopB, SopD, and SopE2 contribute to *Salmonella enterica* serotype Typhimurium invasion of epithelial cells. *Infect Immun* 73: 146-154.
- Raivio, T. L., (2005) Envelope stress responses and Gram-negative bacterial pathogenesis. Mol Microbiol 56: 1119-1128.
- Rang, C., E. Alix, C. Felix, A. Heitz, L. Tasse & A. B. Blanc-Potard, (2007) Dual role of the MgtC virulence factor in host and non-host environments. *Mol Microbiol* **63**: 605-622.
- Rathman, M., L. P. Barker & S. Falkow, (1997) The unique trafficking pattern of *Salmonella* Typhimurium-containing phagosomes in murine macrophages is independent of the mechanism of bacterial entry. *Infect Immun* **65**: 1475-1485.
- Rathman, M., M. D. Sjaastad & S. Falkow, (1996) Acidification of phagosomes containing *Salmonella* Typhimurium in murine macrophages. *Infect Immun* **64**: 2765-2773.
- Resch, A., T. Afonyushkin, T. B. Lombo, K. J. McDowall, U. Blasi & V. R. Kaberdin, (2008) Translational activation by the noncoding RNA DsrA involves alternative RNase III processing in the rpoS 5'-leader. *RNA* **14**: 454-459.
- Rhodius, V. A., W. C. Suh, G. Nonaka, J. West & C. A. Gross, (2006) Conserved and variable functions of the sigmaE stress response in related genomes. *PLoS Biol* 4: e2.
- Romby, P., F. Vandenesch & E. G. Wagner, (2006) The role of RNAs in the regulation of virulence-gene expression. *Curr Opin Microbiol* **9**: 229-236.
- Romling, U., Z. Bian, M. Hammar, W. D. Sierralta & S. Normark, (1998) Curli fibers are highly conserved between *Salmonella* Typhimurium and *Escherichia coli* with respect to operon structure and regulation. *J Bacteriol* **180**: 722-731.
- Rouf, S. F., N. Anwar, M. O. Clements & M. Rhen, (2011) Genetic analysis of the *pnp-deaD* genetic region reveals membrane lipoprotein NlpI as an independent participant in cold acclimatization of *Salmonella enterica* serovar Typhimurium. *FEMS Microbiol Lett* **325**: 56-63.
- Rowley, G., M. Spector, J. Kormanec & M. Roberts, (2006) Pushing the envelope: extracytoplasmic stress responses in bacterial pathogens. *Nat Rev Microbiol* **4**: 383-394.
- Rytkonen, A., J. Poh, J. Garmendia, C. Boyle, A. Thompson, M. Liu, P. Freemont, J. C. Hinton & D. W. Holden, (2007) SseL, a *Salmonella* deubiquitinase required for macrophage killing and virulence. *Proc Natl Acad Sci U S A* **104**: 3502-3507.
- Sabbagh, S. C., C. G. Forest, C. Lepage, J. M. Leclerc & F. Daigle, (2010) So similar, yet so different: uncovering distinctive features in the genomes of *Salmonella enterica* serovars Typhimurium and Typhi. *FEMS Microbiol Lett* **305**: 1-13.
- Saini, S., J. D. Brown, P. D. Aldridge & C. V. Rao, (2008) FliZ Is a posttranslational activator of FlhD4C2-dependent flagellar gene expression. *J Bacteriol* **190**: 4979-4988.
- Saini, S. & C. V. Rao, (2010) SprB is the molecular link between *Salmonella* pathogenicity island 1 (SPI1) and SPI4. *J Bacteriol* **192**: 2459-2462.
- Salcedo, S. P. & D. W. Holden, (2003) SseG, a virulence protein that targets *Salmonella* to the Golgi network. *EMBO J* 22: 5003-5014.
- Saliba, A. E., A. J. Westermann, S. A. Gorski & J. Vogel, (2014) Single-cell RNA-seq: advances and future challenges. *Nucleic Acids Res* **42**: 8845-8860.
- Salmon, K., S. P. Hung, K. Mekjian, P. Baldi, G. W. Hatfield & R. P. Gunsalus, (2003) Global gene expression profiling in *Escherichia coli* K12. The effects of oxygen availability and FNR. *J Biol Chem* **278**: 29837-29855.
- Sambrook, J., Fritsch, E.F and Maniatis, T., (1989) Molecular Cloning: A Laboratory Manual. *Cold Spring Harbour Press*.

- Samuels, D. J., J. G. Frye, S. Porwollik, M. McClelland, J. Mrazek, T. R. Hoover & A. C. Karls, (2013) Use of a promiscuous, constitutively-active bacterial enhancer-binding protein to define the sigma(5)(4) (RpoN) regulon of *Salmonella* Typhimurium LT2. *BMC Genomics* 14: 602
- Santiviago, C. A., M. M. Reynolds, S. Porwollik, S. H. Choi, F. Long, H. L. Andrews-Polymenis & M. McClelland, (2009) Analysis of pools of targeted *Salmonella* deletion mutants identifies novel genes affecting fitness during competitive infection in mice. *PLoS Pathog* 5: e1000477.
- Sauer, E., (2013) Structure and RNA-binding properties of the bacterial LSm protein Hfq. RNA Biol 10: 610-618.
- Schechter, L. M. & C. A. Lee, (2001) AraC/XylS family members, HilC and HilD, directly bind and derepress the *Salmonella* Typhimurium *hilA* promoter. *Mol Microbiol* **40**: 1289-1299.
- Schmitt, C. K., S. C. Darnell & A. D. O'Brien, (1996) The attenuated phenotype of a *Salmonella* Typhimurium *flgM* mutant is related to expression of FliC flagellin. *J Bacteriol* **178**: 2911-2915.
- Schneider, C. A., W. S. Rasband & K. W. Eliceiri, (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**: 671-675.
- Schraidt, O., M. D. Lefebre, M. J. Brunner, W. H. Schmied, A. Schmidt, J. Radics, K. Mechtler, J. E. Galan & T. C. Marlovits, (2010) Topology and organization of the *Salmonella* Typhimurium type III secretion needle complex components. *PLoS Pathog* **6**: e1000824.
- Schraidt, O. & T. C. Marlovits, (2011) Three-dimensional model of *Salmonella*'s needle complex at subnanometer resolution. *Science* **331**: 1192-1195.
- Sengupta, C., S. Ray & R. Chowdhury, (2014) Fine tuning of virulence regulatory pathways in enteric bacteria in response to varying bile and oxygen concentrations in the gastrointestinal tract. *Gut Pathog* **6**: 38.
- Seshasayee, A. S., K. Sivaraman & N. M. Luscombe, (2011) An overview of prokaryotic transcription factors: a summary of function and occurrence in bacterial genomes. Subcell Biochem 52: 7-23.
- Shannon, P., A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski & T. Ideker, (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13: 2498-2504.
- Sharma, C. M., F. Darfeuille, T. H. Plantinga & J. Vogel, (2007) A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites. *Genes Dev* 21: 2804-2817.
- Sharma, C. M., S. Hoffmann, F. Darfeuille, J. Reignier, S. Findeiss, A. Sittka, S. Chabas, K. Reiche, J. Hackermuller, R. Reinhardt, P. F. Stadler & J. Vogel, (2010) The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature* **464**: 250-255.
- Sharma, C. M., K. Papenfort, S. R. Pernitzsch, H. J. Mollenkopf, J. C. Hinton & J. Vogel, (2011) Pervasive post-transcriptional control of genes involved in amino acid metabolism by the Hfq-dependent GcvB small RNA. *Mol Microbiol* 81: 1144-1165.
- Shea, J. E., M. Hensel, C. Gleeson & D. W. Holden, (1996) Identification of a virulence locus encoding a second type III secretion system in *Salmonella* Typhimurium. *Proc Natl Acad Sci U S A* **93**: 2593-2597.
- Shi, Y., T. Latifi, M. J. Cromie & E. A. Groisman, (2004) Transcriptional control of the antimicrobial peptide resistance *ugtL* gene by the *Salmonella* PhoP and SlyA regulatory proteins. *J Biol Chem* **279**: 38618-38625.
- Shimoni, Y., G. Friedlander, G. Hetzroni, G. Niv, S. Altuvia, O. Biham & H. Margalit, (2007) Regulation of gene expression by small non-coding RNAs: a quantitative view. *Mol Syst Biol* 3: 138.
- Shin, D. & E. A. Groisman, (2005) Signal-dependent binding of the response regulators PhoP and PmrA to their target promoters in vivo. *J Biol Chem* **280**: 4089-4094.
- Shotland, Y., H. Kramer & E. A. Groisman, (2003) The *Salmonella* SpiC protein targets the mammalian Hook3 protein function to alter cellular trafficking. *Mol Microbiol* **49**: 1565-1576.

- Silphaduang, U., M. Mascarenhas, M. Karmali & B. K. Coombes, (2007) Repression of intracellular virulence factors in *Salmonella* by the Hha and YdgT nucleoid-associated proteins. *J Bacteriol* **189**: 3669-3673.
- Silva, I. J., A. D. Ortega, S. C. Viegas, F. Garcia-Del Portillo & C. M. Arraiano, (2013) An RpoSdependent sRNA regulates the expression of a chaperone involved in protein folding. *RNA* 19: 1253-1265.
- Singer, H. M., C. Kuhne, J. A. Deditius, K. T. Hughes & M. Erhardt, (2014) The *Salmonella Spi1* virulence regulatory protein HilD directly activates transcription of the flagellar master operon *flhDC*. *J Bacteriol* **196**: 1448-1457.
- Sittka, A., S. Lucchini, K. Papenfort, C. M. Sharma, K. Rolle, T. T. Binnewies, J. C. Hinton & J. Vogel, (2008) Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS Genet* 4: e1000163.
- Sittka, A., V. Pfeiffer, K. Tedin & J. Vogel, (2007) The RNA chaperone Hfq is essential for the virulence of *Salmonella* Typhimurium. *Mol Microbiol* **63**: 193-217.
- Sivasankaran, S. K., (2014) Investigation of the transcriptional landscape and RNA biology of *Salmonella* Typhimurium plasmids. In: School of Genetics and Microbiology. Trinity College Dublin, pp.
- Skinner, M. E., A. V. Uzilov, L. D. Stein, C. J. Mungall & I. H. Holmes, (2009) JBrowse: a next-generation genome browser. *Genome Res* **19**: 1630-1638.
- Skovierova, H., G. Rowley, B. Rezuchova, D. Homerova, C. Lewis, M. Roberts & J. Kormanec, (2006) Identification of the sigmaE regulon of *Salmonella enterica* serovar Typhimurium. *Microbiology* **152**: 1347-1359.
- Sledjeski, D. & S. Gottesman, (1995) A small RNA acts as an antisilencer of the H-NS-silenced rcsA gene of Escherichia coli. Proc Natl Acad Sci U S A 92: 2003-2007.
- Soncini, F. C., E. Garcia Vescovi, F. Solomon & E. A. Groisman, (1996) Molecular basis of the magnesium deprivation response in *Salmonella* Typhimurium: identification of PhoPregulated genes. *J Bacteriol* **178**: 5092-5099.
- Soncini, F. C., E. G. Vescovi & E. A. Groisman, (1995) Transcriptional autoregulation of the *Salmonella* Typhimurium *phoPQ* operon. *J Bacteriol* **177**: 4364-4371.
- Song, H., W. Kong, N. Weatherspoon, G. Qin, W. Tyler, J. Turk, R. Curtiss, 3rd & Y. Shi, (2008) Modulation of the regulatory activity of bacterial two-component systems by SlyA. *J Biol Chem* **283**: 28158-28168.
- Soper, T., P. Mandin, N. Majdalani, S. Gottesman & S. A. Woodson, (2010) Positive regulation by small RNAs and the role of Hfq. *Proc Natl Acad Sci U S A* **107**: 9602-9607.
- Spory, A., A. Bosserhoff, C. von Rhein, W. Goebel & A. Ludwig, (2002) Differential regulation of multiple proteins of *Escherichia coli* and *Salmonella enterica* serovar Typhimurium by the transcriptional regulator SlyA. *J Bacteriol* **184**: 3549-3559.
- Srikumar, S., C. Kröger, M. Hebrard, A. Colgan, M. Beckett, L. Luque, L. Wang, S. Sivasankaran, K. Breen, D. Zhou, A. D. S. Cameron, K. Hokamp & J. C. D. Hinton, (2014) The intra-macrophage transcriptome of *Salmonella* Typhimurium identifies a genus-specific small RNA required for virulence. *PLOS Pathog (in revision)*.
- Stapleton, M. R., V. A. Norte, R. C. Read & J. Green, (2002) Interaction of the *Salmonella* Typhimurium transcription and virulence factor SlyA with target DNA and identification of members of the SlyA regulon. *J Biol Chem* **277**: 17630-17637.
- Stark, T. & R. E. Moses, (1989) Interaction of the LexA repressor and the *uvrC* regulatory region. *FEBS Lett* **258**: 39-41.
- Stender, S., A. Friebel, S. Linder, M. Rohde, S. Mirold & W. D. Hardt, (2000) Identification of SopE2 from *Salmonella* Typhimurium, a conserved guanine nucleotide exchange factor for Cdc42 of the host cell. *Mol Microbiol* **36**: 1206-1221.
- Sterzenbach, T., K. T. Nguyen, S. P. Nuccio, M. G. Winter, C. A. Vakulskas, S. Clegg, T. Romeo & A. J. Baumler, (2013) A novel CsrA titration mechanism regulates fimbrial gene expression in *Salmonella* Typhimurium. *EMBO J* 32: 2872-2883.
- Stoebel, D. M., A. Free & C. J. Dorman, (2008) Anti-silencing: overcoming H-NS-mediated repression of transcription in Gram-negative enteric bacteria. *Microbiology* **154**: 2533-2545.

- Stormo, G. D. & Y. Zhao, (2010) Determining the specificity of protein-DNA interactions. *Nat Rev Genet* 11: 751-760.
- Stringer, A. M., N. Singh, A. Yermakova, B. L. Petrone, J. J. Amarasinghe, L. Reyes-Diaz, N. J. Mantis & J. T. Wade, (2012) FRUIT, a scar-free system for targeted chromosomal mutagenesis, epitope tagging, and promoter replacement in *Escherichia coli* and *Salmonella enterica*. *PLoS One* 7: e44841.
- Sturm, A., M. Heinemann, M. Arnoldini, A. Benecke, M. Ackermann, M. Benz, J. Dormann & W. D. Hardt, (2011) The cost of virulence: retarded growth of *Salmonella* Typhimurium cells expressing type III secretion system 1. *PLoS Pathog* 7: e1002143.
- Teixido, L., B. Carrasco, J. C. Alonso, J. Barbe & S. Campoy, (2011) Fur activates the expression of *Salmonella enterica* pathogenicity island 1 by directly interacting with the *hilD* operator in vivo and in vitro. *PLoS One* **6**: e19711.
- Teplitski, M., R. I. Goodier & B. M. Ahmer, (2003) Pathways leading from BarA/SirA to motility and virulence gene expression in *Salmonella*. *J Bacteriol* **185**: 7257-7265.
- Thiennimitr, P., S. E. Winter & A. J. Baumler, (2012) *Salmonella*, the host and its microbiota. *Curr Opin Microbiol* **15**: 108-114.
- Thijs, I. M., S. C. De Keersmaecker, A. Fadda, K. Engelen, H. Zhao, M. McClelland, K. Marchal & J. Vanderleyden, (2007) Delineation of the *Salmonella enterica* serovar Typhimurium HilA regulon through genome-wide location and transcript analysis. *J Bacteriol* **189**: 4587-4596.
- Thomas-Chollier, M., M. Defrance, A. Medina-Rivera, O. Sand, C. Herrmann, D. Thieffry & J. van Helden, (2011) RSAT 2011: regulatory sequence analysis tools. *Nucleic Acids Res* **39**: W86-91.
- Thomason, M. K., F. Fontaine, N. De Lay & G. Storz, (2012) A small RNA that regulates motility and biofilm formation in response to changes in nutrient availability in *Escherichia coli. Mol Microbiol* **84**: 17-35.
- Tjaden, B., S. S. Goodwin, J. A. Opdyke, M. Guillier, D. X. Fu, S. Gottesman & G. Storz, (2006) Target prediction for small, noncoding RNAs in bacteria. *Nucleic Acids Res* **34**: 2791-2802.
- Tomljenovic-Berube, A. M., D. T. Mulder, M. D. Whiteside, F. S. Brinkman & B. K. Coombes, (2010) Identification of the regulatory logic controlling *Salmonella* pathoadaptation by the SsrA-SsrB two-component system. *PLoS Genet* **6**: e1000875.
- Tran, H. J., A. K. Heroven, L. Winkler, T. Spreter, B. Beatrix & P. Dersch, (2005) Analysis of RovA, a transcriptional regulator of *Yersinia pseudotuberculosis* virulence that acts through antirepression and direct transcriptional activation. *J Biol Chem* 280: 42423-42432.
- Troxell, B., R. C. Fink, S. Porwollik, M. McClelland & H. M. Hassan, (2011a) The Fur regulon in anaerobically grown *Salmonella enterica* sv. Typhimurium: identification of new Fur targets. *BMC Microbiol* 11: 236.
- Troxell, B., M. L. Sikes, R. C. Fink, A. Vazquez-Torres, J. Jones-Carson & H. M. Hassan, (2011b) Fur negatively regulates *hns* and is required for the expression of HilA and virulence in *Salmonella enterica* serovar Typhimurium. *J Bacteriol* **193**: 497-505.
- Tsang, J. & T. R. Hoover, (2014) Themes and Variations: Regulation of RpoN-Dependent Flagellar Genes across Diverse Bacterial Species. *Scientifica (Cairo)* **2014**: 681754.
- Tu, X., T. Latifi, A. Bougdour, S. Gottesman & E. A. Groisman, (2006) The PhoP/PhoQ two-component system stabilizes the alternative sigma factor RpoS in *Salmonella enterica*. *Proc Natl Acad Sci U S A* **103**: 13503-13508.
- Turnbull, A. L. & M. G. Surette, (2010) Cysteine biosynthesis, oxidative stress and antibiotic resistance in *Salmonella* Typhimurium. *Res Microbiol* **161**: 643-650.
- Typas, A., G. Becker & R. Hengge, (2007) The molecular basis of selective promoter activation by the sigmaS subunit of RNA polymerase. *Mol Microbiol* **63**: 1296-1306.
- Urban, J. H. & J. Vogel, (2007) Translational control and target recognition by *Escherichia coli* small RNAs in vivo. *Nucleic Acids Res* **35**: 1018-1037.
- Urban, J. H. & J. Vogel, (2008) Two seemingly homologous noncoding RNAs act hierarchically to activate *glmS* mRNA translation. *PLoS Biol* **6**: e64.

- Uzzau, S., N. Figueroa-Bossi, S. Rubino & L. Bossi, (2001) Epitope tagging of chromosomal genes in *Salmonella*. *Proc Natl Acad Sci U S A* **98**: 15264-15269.
- Vanderpool, C. K. & S. Gottesman, (2004) Involvement of a novel transcriptional activator and small RNA in post-transcriptional regulation of the glucose phosphoenolpyruvate phosphotransferase system. *Mol Microbiol* **54**: 1076-1089.
- Velayudhan, J., M. Castor, A. Richardson, K. L. Main-Hester & F. C. Fang, (2007) The role of ferritins in the physiology of *Salmonella enterica* sv. Typhimurium: a unique role for ferritin B in iron-sulphur cluster repair and virulence. *Mol Microbiol* **63**: 1495-1507.
- Viegas, S. C. & C. M. Arraiano, (2008) Regulating the regulators: How ribonucleases dictate the rules in the control of small non-coding RNAs. *RNA Biol* **5**: 230-243.
- Viegas, S. C., V. Pfeiffer, A. Sittka, I. J. Silva, J. Vogel & C. M. Arraiano, (2007) Characterization of the role of ribonucleases in *Salmonella* small RNA decay. *Nucleic Acids Res* 35: 7651-7664.
- Vivero, A., R. C. Banos, J. F. Mariscotti, J. C. Oliveros, F. Garcia-del Portillo, A. Juarez & C. Madrid, (2008) Modulation of horizontally acquired genes by the Hha-YdgT proteins in *Salmonella enterica* serovar Typhimurium. *J Bacteriol* **190**: 1152-1156.
- Vogel, J., (2009) A rough guide to the non-coding RNA world of *Salmonella*. *Mol Microbiol* **71**: 1-11.
- Vogel, J., L. Argaman, E. G. Wagner & S. Altuvia, (2004) The small RNA IstR inhibits synthesis of an SOS-induced toxic peptide. *Curr Biol* **14**: 2271-2276.
- Vogel, J., V. Bartels, T. H. Tang, G. Churakov, J. G. Slagter-Jager, A. Huttenhofer & E. G. Wagner, (2003) RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res* 31: 6435-6443.
- Vogel, J. & B. F. Luisi, (2011) Hfq and its constellation of RNA. Nat Rev Microbiol 9: 578-589.
- Vogel, J. & E. G. Wagner, (2007) Target identification of small noncoding RNAs in bacteria. *Curr Opin Microbiol* **10**: 262-270.
- Wada, T., Y. Tanabe & K. Kutsukake, (2011) FliZ acts as a repressor of the *ydiV* gene, which encodes an anti-FlhD4C2 factor of the flagellar regulon in *Salmonella enterica* serovar Typhimurium. *J Bacteriol* **193**: 5191-5198.
- Wagner, C. & M. Hensel, (2011) Adhesive mechanisms of *Salmonella enterica*. *Adv Exp Med Biol* **715**: 17-34.
- Wagner, E. G., (2013) Cycling of RNAs on Hfq. RNA Biol 10: 619-626.
- Wagner, G. P., K. Kin & V. J. Lynch, (2012) Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci* **131**: 281-285.
- Wagner, G. P., K. Kin & V. J. Lynch, (2013) A model based criterion for gene expression calls using RNA-seq data. *Theory Biosci* **132**: 159-164.
- Walthers, D., R. K. Carroll, W. W. Navarre, S. J. Libby, F. C. Fang & L. J. Kenney, (2007) The response regulator SsrB activates expression of diverse *Salmonella* pathogenicity island 2 promoters and counters silencing by the nucleoid-associated protein H-NS. *Mol Microbiol* 65: 477-493.
- Wanner, B. L., (1996) Phosphorous assimilation and control of the phosphate regulon in *Escherichia coli* and *Salmonella* Typhimurium. In: Cellular and Molecular Biology. ASM Press, pp. 1357-1381.
- Wilson, D., V. Charoensawan, S. K. Kummerfeld & S. A. Teichmann, (2008) DBD-taxonomically broad transcription factor predictions: new content and functionality. *Nucleic Acids Res* **36**: D88-92.
- Winter, S. E., P. Thiennimitr, M. G. Winter, B. P. Butler, D. L. Huseby, R. W. Crawford, J. M. Russell, C. L. Bevins, L. G. Adams, R. M. Tsolis, J. R. Roth & A. J. Baumler, (2010) Gut inflammation provides a respiratory electron acceptor for *Salmonella*. *Nature* 467: 426-429.
- Wray, C. & W. J. Sojka, (1978) Experimental *Salmonella* Typhimurium infection in calves. *Res Vet Sci* **25**: 139-143.
- Wright, P. R., J. Georg, M. Mann, D. A. Sorescu, A. S. Richter, S. Lott, R. Kleinkauf, W. R. Hess & R. Backofen, (2014) CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains. *Nucleic Acids Res* **42**: W119-123.

- Wright, P. R., A. S. Richter, K. Papenfort, M. Mann, J. Vogel, W. R. Hess, R. Backofen & J. Georg, (2013) Comparative genomics boosts target prediction for bacterial small RNAs. *Proc Natl Acad Sci U S A* **110**: E3487-3496.
- Xu, X. & M. Hensel, (2010) Systematic analysis of the SsrAB virulon of *Salmonella enterica*. *Infect Immun* **78**: 49-58.
- Ygberg, S. E., M. O. Clements, A. Rytkonen, A. Thompson, D. W. Holden, J. C. Hinton & M. Rhen, (2006) Polynucleotide phosphorylase negatively controls *spv* virulence gene expression in *Salmonella enterica*. *Infect Immun* 74: 1243-1254.
- Yoon, H., C. Ansong, J. E. McDermott, M. Gritsenko, R. D. Smith, F. Heffron & J. N. Adkins, (2011) Systems analysis of multiple regulator perturbations allows discovery of virulence factors in *Salmonella*. *BMC Syst Biol* 5: 100.
- Yoon, H., J. E. McDermott, S. Porwollik, M. McClelland & F. Heffron, (2009) Coordinated regulation of virulence during systemic infection of *Salmonella enterica* serovar Typhimurium. *PLoS Pathog* **5**: e1000306.
- Yu, J. & T. Schneiders, (2012) Tigecycline challenge triggers sRNA production in *Salmonella* enterica serovar Typhimurium. *BMC Microbiol* 12: 195.
- Zadeh, J. N., C. D. Steenberg, J. S. Bois, B. R. Wolfe, M. B. Pierce, A. R. Khan, R. M. Dirks & N. A. Pierce, (2011) NUPACK: Analysis and design of nucleic acid systems. *J Comput Chem* **32**: 170-173.
- Zhang, A., K. M. Wassarman, C. Rosenow, B. C. Tjaden, G. Storz & S. Gottesman, (2003) Global analysis of small RNA and mRNA targets of Hfq. *Mol Microbiol* **50**: 1111-1124.
- Zhou, D. & J. Galan, (2001) *Salmonella* entry into host cells: the work in concert of type III secreted effector proteins. *Microbes Infect* 3: 1293-1298.
- Zhou, D., M. S. Mooseker & J. E. Galan, (1999) Role of the *S.* typhimurium actin-binding protein SipA in bacterial internalization. *Science* **283**: 2092-2095.
- Zwir, I., T. Latifi, J. C. Perez, H. Huang & E. A. Groisman, (2012) The promoter architectural landscape of the *Salmonella* PhoP regulon. *Mol Microbiol* **84**: 463-485.