



---

# Transcriptomic investigation of the adaptation of *Streptococcus pneumoniae*

---

**Modupeh Betts**

A thesis submitted for the degree of  
**Doctor of Philosophy**  
University College London

September 2022

**Declaration**

I, Modupeh Betts, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Modupeh Betts,

London, September 2022

## Abstract

*Streptococcus pneumoniae* colonises the human nasopharynx as a commensal but can translocate to the lungs, meninges, and blood to cause potentially fatal infections. These host niches exhibit diverse physiological environments. Differences in adaptation to these conditions may explain differences between serotypes and genotypes in their ability to colonise the human host, be transmitted, and to cause disease. RNA sequencing (RNA-Seq) was used to investigate adaptation of clinical *S. pneumoniae* strains to different stress environments. In Chapter 3, to establish the optimal experimental conditions, the effects of carbohydrate source, temperature, and iron concentrations on bacterial growth dynamics were evaluated. *S. pneumoniae* strains selected on the basis of their ability to be carried and cause disease, showed differential growth phenotypes. In Chapter 4, to facilitate robust transcriptomic analysis, high-quality genome assemblies of *S. pneumoniae* serotype 1 (highly invasive, rarely found in carriage) and serotype 6B (rarely invasive, highly carried) strains were generated and characterised. A pneumococcal transcriptomic analysis pipeline was developed in Chapter 5 by investigating the transcriptomic response of two single gene knockouts of *S. pneumoniae* serotype 6B lacking the biosynthesis genes *fhs* or *proABC*. These mutants have been shown to be attenuated *in vivo* and the aim was to identify the transcriptomic basis for this. Adaptation by  $\Delta fhs$  *S. pneumoniae* included upregulation of pathways involved in secondary metabolites biosynthesis and quorum sensing while the  $\Delta proABC$  *S. pneumoniae* was upregulated for carbohydrate metabolism pathways. In Chapters 6 and 7, the transcriptomic adaptations of *S. pneumoniae* serotype 1 and serotype 6B strains to altered iron and temperature levels were delineated respectively, indicating strain specific gene expression with the majority of differential regulation occurring in core pneumococcal genes. In Chapter 8, to pave the way for investigating the *S. pneumoniae* transcriptome in human samples, a challenge in pneumococcal research, an approach to directly isolate high-quality pneumococcal RNA from human carriers was developed. The work in this thesis provides new insights in the gene regulation of clinical *S. pneumoniae* strains under various environmental exposures.

## Impact Statement

*Streptococcus pneumoniae* is a leading cause of pneumonia, meningitis, and sepsis globally and has been designated a priority pathogen for research and development by the World Health Organisation (WHO). Although asymptomatic carriage of *S. pneumoniae* by humans is considered a prerequisite for disease development, the mechanisms that govern the transformation of a commensal to cause potentially fatal disease remains elusive. The work in this thesis makes original contributions towards our understanding of this process by leveraging next generation sequencing methods to investigate gene expression of *S. pneumoniae* strains. We have established experimental and analysis methodologies for conducting comparative transcriptomics analysis of clinically important *S. pneumoniae* strains. RNA-Seq was used to compare the adaptation of a “highly invasive” and “relatively less invasive” *S. pneumoniae* strains, providing novel insights into pneumococcal biology and how differential adaptation of strains may influence bacterial carriage, transmission, and disease development. The outcomes of this work not only improve understanding of pneumococcal pathogenesis but highlight potential vaccine targets. The work is highly relevant to the study of bacterial pathogenesis in general. A key strength of the work undertaken in this thesis is the use of clinically important strains, deviating from previous experimental studies that use laboratory adapted strains which have limited generalisability. We have demonstrated here that the adaptation of strains with different invasive potential is associated with differential expression of core pneumococcal genes. This work has also contributed complete genome assemblies of two clinically important *S. pneumoniae* strains, providing a valuable resource to the wider scientific community. Methods developed in this thesis may lead to the investigation of the *S. pneumoniae* transcriptome directly from human samples, a current challenge in pneumococcal research. In conclusion, this work has established methodologies for comparative transcriptomic analysis of *S. pneumoniae*, provided new insights into biology of the pathogen, and provided the scientific community with a novel set of high throughput datasets for further elucidation of *S. pneumoniae* pathogenesis.



## **Acknowledgements**

I wish to express my immense gratitude to my supervisor Prof. Robert Heyderman for his scientific guidance, kindness, and unrelenting support to enable me to complete this work. I am extremely lucky to have been able to count on your leadership and wisdom during this phase of my academic development - a terrific mentor but also an exceptional human being.

I would like to thank my thesis committee members Prof. Jeremy Brown, Prof. Francois Balloux, and Dr Kristine Arnvig for helping to shape my research ideas and for providing helpful feedback on the design of the project. I am grateful to my secondary supervisor Prof. Jeremy Brown for accepting to be part of this work, always being excited to discuss data and providing critical feedbacks. I am thankful to Prof. Francois Balloux for allowing me to be part of his weekly laboratory meetings discussing microbial genomics. Dr Kristine Arnvig's expertise in bacterial gene regulation was instrumental in determining transcriptomics experiments to take forward in this work.

I am especially thankful to members of the Heyderman group at UCL for contributing to a conducive camaraderie work environment— Dr Andrea Gori, Dr Caroline Weight, Dr Elizabeth Chan, Dr Brenda Kwambana-Adams, Dr Asia-Sophia Wolf, Dr Lavanya Mane, and Dr Akuzike Kalizang'oma. I want to thank the entire team for helpful discussions of my data during the weekly laboratory meetings. Thank you, Dr Caroline Weight, for inducting me into the laboratory and for always willing to review my written work and returning in record time. Thank you also for the numerous impromptu uplifting chats. Thank you, Dr Andrea Gori and Dr Elizabeth Chan, for helpful suggestions on my data. Many thanks to Dr Lavanya Mane for training me to use the 7H9+ media she developed as part of her PhD. Thanks to my sister Dr Brenda Kwambana-Adams who has been long involved in my academic journey and remains an inspiration. I am thankful to Dr Emma Wall for discussions on pneumococcal RNA isolation methodologies and for sharing her protocol which was adapted and used in this thesis.

I want to thank MPRU administrators Steven Kuo and Dr Anna Wang-Erickson for the tremendous support they provided me along the way. My appreciation extends to all supports staff at Rayne building UCL for making it easy to focus on our research. I want to thank Carolyn McElvaney, divisional finance manager, for her prompt responses to my many enquiries and always helping to find solutions to my problems. I am also thankful to NIHR for providing the funds that supported my PhD project.

I wish to thank Prof. Jay Hinton and members of his team at the University of Liverpool for the support they rendered to this work. Many thanks to Dr Blanca Perez-Sepulveda and Dr Rocio Canals for laboratory training and discussions on *Salmonella* transcriptomics experiments which I adapted for *S. pneumoniae* here. I am thankful to Dr Alex Predeus for important discussions on bioinformatics analysis of my transcriptomics dataset and for sharing his transcriptomics analysis pipeline. Thanks to Yan Li for generating the Jbrowse database of our transcriptomic data for the iron experiment.

I wish to also thank Prof. Daniela Ferreira and her team at Liverpool School of Tropical Medicine for providing samples from the experimental human challenge model to include in this work.

I would not have been in a position to undertake a PhD without the prior support, directorship, and encouragement of Prof. Martin Antonio, Prof. Tumani Corrah, Prof. Andrew Prentice, and Dr Matt Silver. I thank you all for your guidance and immense contributions to my academic development.

Finally, I would like to thank my entire family for their support. I am especially thankful for the love and care of my mum, Laliya Yabou Hydara. Thank you, mum, for your unflinching support and plentiful prayers during this journey. I thank my dad Maulana M Betts, my sister Aissatou Betts, and my brother Mamina Betts for the encouragement and support. I am equally thankful to my cousin Ebou Mbaye. During the most difficult times of this journey, I found solace and inspiration through my son.

## Research Paper Declaration Form:

<b>1. For a research manuscript that has already been published</b> (if not yet published, please skip to section 2):		
<b>a) Where was the work published?</b> (e.g. journal name)	Microbiology Resource Announcements	
<b>b) Who published the work?</b> (e.g. Elsevier/Oxford University Press):	American Society for Microbiology Journals	
<b>c) When was the work published?</b>	30/09/2021	
<b>d) Was the work subject to academic peer review?</b>		
<b>e) Have you retained the copyright for the work?</b>		
[If no, please seek permission from the relevant publisher and check the box next to the below statement]: <i>I acknowledge permission of the publisher named under 1b to include in this thesis portions of the publication named as included in 1a.</i>		
<b>2. For a research manuscript prepared for publication but that has not yet been published</b> (if already published, please skip to section 3):		
<b>a) Has the manuscript been uploaded to a preprint server?</b> (e.g. medRxiv):	Please select.	<b>If yes, which server?</b> Click or tap here to enter text.
<b>b) Where is the work intended to be published?</b> (e.g. names of journals that you are planning to submit to)	Click or tap here to enter text.	
<b>c) List the manuscript's authors in the intended authorship order:</b>	Click or tap here to enter text.	
<b>d) Stage of publication</b>	Please select.	
<b>3. For multi-authored work, please give a statement of contribution covering all authors</b> (if single-author, please skip to section 4):		
<p>Modupeh Betts: Conceptualization, Data curation, Formal analysis, Funding, acquisition, Investigation, Methodology, Project administration, Resources, Validation, Visualization, Writing - original draft, Writing - review &amp; editing</p> <p>Seth Jarvis: Funding acquisition, Writing - review &amp; editing</p> <p>Aaron Jeffries: Data curation, Formal analysis, Writing - review &amp; editing</p> <p>Andrea Gori: Supervision, Writing - review &amp; editing</p> <p>Chrispin Chaguzza: Supervision, Writing - review &amp; editing</p> <p>Jacqueline Msefula: Data curation, Methodology</p> <p>Brenda Kwambana-Adams: Supervision, Writing - review &amp; editing</p> <p>Neil French: Data curation</p>		

<p>Todd Swarthout: Data curation, Methodology, Writing - review &amp; editing</p> <p>Jeremy S Brown: Supervision, Writing - review &amp; editing</p> <p>Robert S Heyderman: Conceptualization, Data curation, Funding acquisition, Project administration, Resources, Supervision, Writing - review &amp; editing</p>			
<p><b>4. In which chapter(s) of your thesis can this material be found?</b></p>			
<p>Chapter 4</p>			
<p><b>5. e-Signatures confirming that the information above is accurate</b> (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work):</p>			
<p><b>Candidate:</b></p>	<p>Modupeh Betts</p>	<p><b>Date:</b></p>	<p>18/09/2022</p>
<p><b>Supervisor/ Senior Author</b> (where appropriate):</p>	<p>Prof. Robert S Heyderman</p>	<p><b>Date:</b></p>	<p>18/09/2022</p>

## Table of Contents

<b>Declaration.....</b>	<b>2</b>
<b>Abstract.....</b>	<b>3</b>
<b>Impact Statement.....</b>	<b>4</b>
<b>Acknowledgements .....</b>	<b>5</b>
<b>Research Paper Declaration Form:.....</b>	<b>7</b>
<b>Table of Contents.....</b>	<b>9</b>
<b>Abbreviations .....</b>	<b>16</b>
<b>Chapter 1 : General introduction.....</b>	<b>19</b>
1.1 Brief history of <i>Streptococcus pneumoniae</i> and its role in modern genetics .....	20
1.2 The epidemiology of <i>Streptococcus pneumoniae</i> .....	20
1.3 <i>S. pneumoniae</i> host niches .....	22
1.4 The nasopharyngeal environment and <i>S. pneumoniae</i> .....	23
1.4.1 Temperature.....	24
1.4.2 Carbohydrate source .....	24
1.4.3 Transition metals .....	24
1.4.4 Other physiological properties of the nasopharyngeal niche.....	25
1.5 Adaptation and virulence factors .....	26
1.6 Microarrays .....	28
1.7 RNA sequencing (RNA-Seq).....	28
1.8 RNA-Seq for bacterial transcriptomics studies .....	29
1.9 Using RNA-Seq to study <i>S. pneumoniae</i> transcriptomics adaptation .....	30
1.10 Thesis aims .....	32
<b>Chapter 2 : General materials and methods.....</b>	<b>34</b>
2.1 Chemicals and reagents .....	35
2.2 General safety and control measures.....	35
2.3 <i>S. pneumoniae</i> strains and rationale for selection.....	35

<b>2.4 Bacterial stocks preparation .....</b>	<b>36</b>
<b>2.5 Isolation and identification of <i>S. pneumoniae</i> serotype 1 (strain BVJ1JL).....</b>	<b>37</b>
2.5.1 Colony forming unit of cultures .....	37
<b>2.6 Growth media.....</b>	<b>38</b>
2.6.1 Todd Hewitt broth + Yeast (enriched media) .....	38
2.6.2 7H9+ medium .....	38
2.6.3 Nose mimicking conditions (NMC) medium .....	38
2.6.4 Lung mimicking conditions (LMC) medium.....	39
2.6.5 Chang CDM .....	39
<b>2.7 Growth curve assays .....</b>	<b>39</b>
<b>2.8 Growth conditions .....</b>	<b>40</b>
<b>2.9 Growth parameters analysis .....</b>	<b>40</b>
<b>2.10 RNA isolation protocol .....</b>	<b>40</b>
<b>2.11 RNA quantification.....</b>	<b>46</b>
<b>2.12 Quality control, mapping and counting RNA-Seq reads .....</b>	<b>47</b>
<b>2.13 Differential gene expression analyses .....</b>	<b>47</b>
<b>2.14 Biological pathways analysis .....</b>	<b>48</b>
<b>2.15 Statistical analysis.....</b>	<b>48</b>
<b>Chapter 3 : Growth curves and optimisation experiments.....</b>	<b>49</b>
<b>3.1 Introduction.....</b>	<b>50</b>
<b>3.2 Results.....</b>	<b>51</b>
3.2.1 Optimisation experiments .....	51
3.2.2 Comparison of growth of <i>S. pneumoniae</i> strains in different growth conditions.....	58
3.2.3 Assessment of RNA isolation protocol for high-quality <i>S. pneumoniae</i> RNA.....	66
<b>3.3 Discussion .....</b>	<b>69</b>
<b>Chapter 4 : Generating high-quality complete genome assemblies of <i>Streptococcus pneumoniae</i> serotype 1 (BVJ1JL) and serotype 6B (BHN418) strains .....</b>	<b>72</b>
<b>4.1 Introduction.....</b>	<b>73</b>
<b>4.2 Methods .....</b>	<b>75</b>
4.2.1 Bacterial strains .....	75

4.2.2 Microbiology culture of <i>S. pneumoniae</i> .....	75
4.2.3 Genomic DNA isolation.....	75
4.2.4 PacBio long read sequencing.....	76
4.2.5 Illumina short read sequencing .....	77
4.2.6 Sequence reads quality control .....	77
4.2.7 Genome assembly and quality assessment .....	77
4.2.8 Genome annotation and curation .....	78
4.2.9 Genetic characterisation of BVJ1JL (serotype 1) and BHN418UCL (serotype 6B).....	78
4.2.10 Comparative genomics and pangenome analysis.....	78
<b>4.3 Results.....</b>	<b>82</b>
4.3.1 Genomic DNA quantity and quality .....	82
4.3.2 PacBio reads sequencing metrics .....	82
4.3.3 Illumina sequencing reads metrics .....	85
4.3.4 HGAP <i>de novo</i> assembly metrics .....	85
4.3.5 Hybrid genome assembly of BHN418UCL (serotype 6B) and BVJ1JL (serotype 1) <i>S. pneumoniae</i> .....	86
4.3.6 Genome annotation and pangenome analysis.....	88
4.3.7 Genetic characterisation of genome assemblies.....	89
4.3.8 Prophage detection .....	90
<b>4.4 Discussion.....</b>	<b>91</b>
<b><i>Chapter 5 : Assessing the stress response of S. pneumoniae mutants lacking the biosynthesis genes fhs and proABC .....</i></b>	<b><i>94</i></b>
<b>5.1 Introduction .....</b>	<b>95</b>
<b>5.2 Methods .....</b>	<b>98</b>
5.2.1 Mutant genes information .....	98
5.2.2 Mutant construction.....	100
5.2.3 Bacterial growth and transcriptomic experiment design .....	100
5.2.4 Library preparation and sequencing.....	100
5.2.5 Quality control, mapping, and quantification of transcripts .....	101
5.2.6 Differential gene expression analyses .....	101
5.2.7 Pathway enrichment analysis .....	102
5.2.8 Identification of correlated gene modules .....	102
5.2.9 Statistical analyses of genesets .....	103
<b>5.3 Results.....</b>	<b>103</b>
5.3.1 <i>S. pneumoniae</i> transcriptome sequencing and mapping of reads .....	103

5.3.2 $\Delta fhs$ and $\Delta proABC$ <i>S. pneumoniae</i> gene knockout verification .....	104
5.3.3 $\Delta fhs$ and $\Delta proABC$ <i>S. pneumoniae</i> transcriptome data quality .....	105
5.3.4 Modest differences in the transcriptome of $\Delta fhs$ and $\Delta proABC$ <i>S. pneumoniae</i> compared to the WT in THY .....	107
5.3.5 $\Delta fhs$ and $\Delta proABC$ <i>S. pneumoniae</i> upregulate expression of fatty-acid synthesis genes in THY .....	109
5.3.6 $\Delta fhs$ and $\Delta proABC$ <i>S. pneumoniae</i> upregulate expression of metabolic genes in THY .....	110
5.3.7 Growth of $\Delta fhs$ and $\Delta proABC$ <i>S. pneumoniae</i> in human serum to induce stress results in a distinct virulence gene expression pattern compared to growth in THY .....	111
5.3.8 Deletion of <i>fhs</i> or <i>proABC</i> genes linked to dysregulation of fatty-acid biosynthesis in <i>S. pneumoniae</i> in serum .....	113
5.3.9 Virulence genes expression by $\Delta fhs$ <i>S. pneumoniae</i> in human serum .....	114
5.3.10 Virulence gene expression by $\Delta proABC$ <i>S. pneumoniae</i> in human serum .....	116
5.3.11 KEGG pathways enrichment analysis for $\Delta fhs$ and $\Delta proABC$ <i>S. pneumoniae</i> in human serum .....	118
5.3.12 Gene co-expression and module identification .....	126
5.3.13 Genes in selected pathways were up regulated in serum but not THY for mutant strains. ....	131
<b>5.4 Discussion .....</b>	<b>134</b>
<b>Chapter 6 : Transcriptomic analysis of the adaptation of <i>S. pneumoniae</i> under Iron-deplete and Iron-replete conditions.....</b>	<b>139</b>
<b>6.1 Introduction.....</b>	<b>140</b>
<b>6.2 Methods .....</b>	<b>146</b>
6.2.1 Bacterial strains and growth media .....	146
6.2.2 Induction of iron-deplete and iron-replete conditions .....	147
6.2.3 Effect of iron depletion on <i>S. pneumoniae</i> growth.....	147
6.2.4 Cellular viability of <i>S. pneumoniae</i> under Iron-deplete conditions.....	148
6.2.5 <i>S. pneumoniae</i> transcriptome in iron-deplete and iron-replete conditions .....	148
6.2.6 RNA isolation and quantification .....	149
6.2.7 Whole transcriptome cDNA .....	151
6.2.8 cDNA from +/- terminator exonuclease (TEX) treatment of RNA (dRNA-Seq) .....	151
6.2.9 Pooled depletion of ribosomal RNA and cDNA size fractionation .....	151
6.2.10 Transcripts mapping and counting .....	152
6.2.11 Differential gene expression analysis .....	152
6.2.12 Statistics and visualization .....	153
6.2.13 Orthologous genes analysis .....	154
6.2.14 Pathway analysis.....	154



6.2.15 Creation of genome browser of reads.....	154
<b>6.3 Results.....</b>	<b>155</b>
6.3.1 Effects of iron-deplete and iron-replete conditions on <i>S. pneumoniae</i> growth .....	155
6.3.2 Assessing potential toxic effects of 2,2' bipyridyl on <i>S. pneumoniae</i> viability.....	158
6.3.3 High yield RNA quantities generated.....	159
6.3.4 Total transcriptomics reads .....	160
6.3.5 RNA-Seq data quality.....	166
6.3.6 Analysis of <i>S. pneumoniae</i> orthologs.....	167
6.3.7 Impact of growth media on <i>S. pneumoniae</i> transcriptome.....	167
6.3.8 Early transcriptomic response of <i>S. pneumoniae</i> during iron limitation in THY .....	170
6.3.9 Later transcriptomics response of <i>S. pneumoniae</i> during iron limitation in THY .....	177
6.3.10 Enriched biological pathways by <i>S. pneumoniae</i> in iron-deplete conditions in THY .....	185
6.3.11 Early transcriptomic response of <i>S. pneumoniae</i> during iron limitation in CDM .....	187
6.3.12 Later transcriptomic response of <i>S. pneumoniae</i> during iron limitation in CDM .....	194
6.3.13 Early transcriptomics response of <i>S. pneumoniae</i> in iron replete CDM .....	200
6.3.14 Later transcriptomic response of <i>S. pneumoniae</i> in iron replete media .....	206
6.3.15 Orthologous genes in <i>S. pneumoniae</i> strains .....	212
6.3.16 Regulated pathways by <i>S. pneumoniae</i> during early adaptation to iron-deplete and iron-replete conditions in CDM .....	213
6.3.17 Regulated pathways by <i>S. pneumoniae</i> during late adaptation to iron-deplete and iron-replete conditions in CDM .....	216
6.3.18 Strain specific transcriptomics changes during growth in iron-deplete and iron-replete conditions .....	218
6.3.19 Analysis of <i>S. pneumoniae</i> orthologs confirmed strain-specific transcriptional adaptation to iron-deplete and iron-replete conditions .....	221
<b>6.4 Discussion.....</b>	<b>227</b>
<b>Chapter 7 : Transcriptomic analysis of the adaptation of <i>S. pneumoniae</i> to changes in temperature .....</b>	<b>233</b>
<b>7.1 Introduction .....</b>	<b>234</b>
<b>7.2 Methods.....</b>	<b>237</b>
7.2.1 Conservation of temperature response genes in clinical <i>S. pneumoniae</i> strains .....	237
7.2.2 <i>S. pneumoniae</i> genomes used for gene conservation analysis .....	238
7.2.3 Investigation of gene conservation .....	239
7.2.4 Bacterial strains for transcriptomics experiment .....	240
7.2.5 Growth media.....	240
7.2.6 Growth experiments.....	240

7.2.7 Temperature changes transcriptomic experiment design.....	241
7.2.8 RNA isolation and quantification .....	243
7.2.9 Library preparation and sequencing.....	243
7.2.10 Sequencing reads quality control, and mapping, transcript counting .....	244
7.2.11 Differential gene expression analysis .....	244
7.2.12 Statistical analysis .....	245
7.2.13 Orthologs and pathway analyses.....	245
<b>7.3 Results.....</b>	<b>246</b>
7.3.1 Differences in growth of <i>S. pneumoniae</i> strains at different temperatures was not explained by genetic differences in temperature-responsive genes .....	246
7.3.2 Growth characteristics of <i>S. pneumoniae</i> at 30°C .....	249
7.3.3 RNA quantification.....	251
7.3.4 Total sequencing reads and mapping .....	251
7.3.5 RNA-Seq data quality.....	255
7.3.6 Transcriptomics response of <i>S. pneumoniae</i> during temperature change from 30°C to 37°C for 10 minutes .....	257
7.3.7 Biological pathways enriched by <i>S. pneumoniae</i> during temperature shift from 30°C to 37°C for 10 minutes .....	267
7.3.8 Transcriptomics response of <i>S. pneumoniae</i> during temperature shift from 30°C to 37°C for 20 minutes.....	269
7.3.9 Biological pathways enriched by <i>S. pneumoniae</i> during temperature shift from 30°C to 37°C for 20 minutes .....	280
7.3.10 Comparison of transcriptomes of <i>S. pneumoniae</i> strains at all timepoints following growth at elevated temperature.....	283
7.3.11 Transcriptomics profiles of selected <i>S. pneumoniae</i> genes in response to temperature change .....	286
<b>7.4 Discussion.....</b>	<b>291</b>
<b>Chapter 8 : Development of techniques for isolation of <i>S. pneumoniae</i> RNA from human samples for transcriptomic analysis .....</b>	<b>297</b>
<b>8.1 Introduction.....</b>	<b>298</b>
<b>8.2 Methods .....</b>	<b>301</b>
8.2.1 Experimental human pneumococcal carriage (EHPC) samples .....	301
8.2.2 RNA isolation of nasopharyngeal samples (NPS) retrieved from healthy volunteers.....	301
8.2.3 RNA extraction of EHPC samples .....	302
8.2.4 RNA quantification.....	305

8.2.5 Confirmation of <i>S. pneumoniae</i> by qPCR .....	306
8.2.6 Data analysis .....	308
8.2.7 Gel electrophoresis .....	309
<b>8.3 Results.....</b>	<b>309</b>
8.3.1 Bacterial quantities retrieved from EHPC participants .....	309
8.3.2 RNA extraction results: EHPC samples batch 1.....	310
8.3.3 RNA extraction results: volunteer samples .....	311
8.3.4 RNA extraction results: EHPC samples batch 2.....	312
8.3.5 RNA extraction results: EHPC samples batch 3.....	312
8.3.6 Modifications to RNA isolation protocol improved RNA quality of EHPC samples.....	313
8.3.7 <i>S. pneumoniae</i> detection in EHPC RNA by quantitative PCR targeting <i>lytA</i> .....	314
8.3.8 <i>S. pneumoniae</i> detection in EHPC RNA samples by quantitative PCR targeting <i>ply</i> .....	315
8.3.9 Confirmation of <i>S. pneumoniae</i> serotype 6B by PCR.....	316
<b>8.4 Discussion.....</b>	<b>318</b>
<b>Chapter 9 : General discussion .....</b>	<b>321</b>
9.1 Summary and general discussion .....	322
9.2 Future Work .....	331
9.3 Limitations.....	333
9.4 Outlook .....	334
9.5 Closing remarks .....	335
<b>Chapter 10 : Appendices .....</b>	<b>336</b>
10.1 Appendix 1: Front page of publication resulting from work in Chapter 4. ....	337
10.2 Appendix 2: Detailed preparation of 7H9+ growth medium .....	338
10.3 Appendix 3: Detailed preparation of NMC growth medium (Aprianto <i>et al.</i> 2016).....	340
<b>Reference list.....</b>	<b>344</b>

## Abbreviations

ABC	ATP-binding cassette
ATP	Adenosine triphosphate
BA	Blood agar
BCAA	Branched chain amino acid
BHI	Brain heart infusion media
BIP	2,2'-Bipyridine/ 2,2'-Bipyridyl/ 2,2'-Dipyridine/ 2,2'-Dipyridyl
BLAST	Basic Local Alignment Search Tool
Bp	base pair
BSA	Bovine serum albumin
CBA	Columbia blood agar
CDM	Chemically defined medium
CFU	Colony forming units
CSP	Competence-stimulating peptide
D39	<i>Streptococcus pneumoniae</i> D39 strain (serotype 2)
DE	Differentially expressed
DEG	Differentially expressed genes
DGE	Differential gene expression
DNA	Deoxyribonucleic acid
e.g.,	For example
EHPC	Experimental human pneumococcal carriage
FAS	Fatty-acid synthesis
FC	Fold change
FDR	False discovery rate
Fe	Iron
G+C	Guanine + cytosine
GWAS	Genome-wide association study
Hrs	hours
HSP	heat shock protein
IGV	Integrative genome visualiser
IPD	invasive pneumococcal disease

Kb	kilobase
KEGG	Kyoto Encyclopaedia of Genes and Genomes
LMC	Lung-mimicking condition medium
LSTM	Liverpool School of Tropical Medicine
MEP	Mid exponential phase
mM	Millimolar
Mn	Manganese
NA	Not applicable
<i>N. meningitides</i>	<i>Neisseria meningitides</i>
NCBI	The National Centre for Biotechnology Information
ncRNA	non-coding ribonucleic acid
NFS	Nasal fluid sample
NMC	Nose-mimicking condition medium
OD <sub>600</sub>	Optical density at 600 nm
OPS	oropharyngeal sample
ORA	Over representation analysis
PBS	Phosphate buffered saline
PCA	Principal component analysis
PCV	Pneumococcal conjugate vaccine
PCV13	13-valent PCV (Prevnar 13)
PCR	Polymerase chain reaction
<i>piu</i>	pneumococcal iron uptake gene
<i>ply</i>	<i>pneumolysin</i> gene
PTS	Phosphotransferase system
QC	Quality control
RBS	Ribosomal binding site
RNA	Ribonucleic acid
RNAT	RNA thermosensor
RNA-Seq	RNA sequencing
ROI	Reactive oxygen intermediates
SEM	Standard error from the mean
<i>S. pneumoniae</i>	<i>Streptococcus pneumoniae</i>

sRNA	small ribonucleic acid
STGG	Skim milk-tryptone-glucose-glycerol
TCS	Two component system
THB	Todd-Hewitt broth
THY	Todd-Hewitt broth + 0.5% yeast extract media
TIGR4	<i>S. pneumoniae</i> TIGR4 strain
TPM	Transcripts per million
TSS	Transcript start sites
TTS	Transcript termination sites
tRNA	Transfer RNA
UCL	University College London
URT	Upper respiratory tract
vs.	Versus
WGS	Whole genome sequencing
WHO	World Health Organisation
WT	Wild type
Zn	Zinc

# Chapter 1 : General introduction

### **1.1 Brief history of *Streptococcus pneumoniae* and its role in modern genetics**

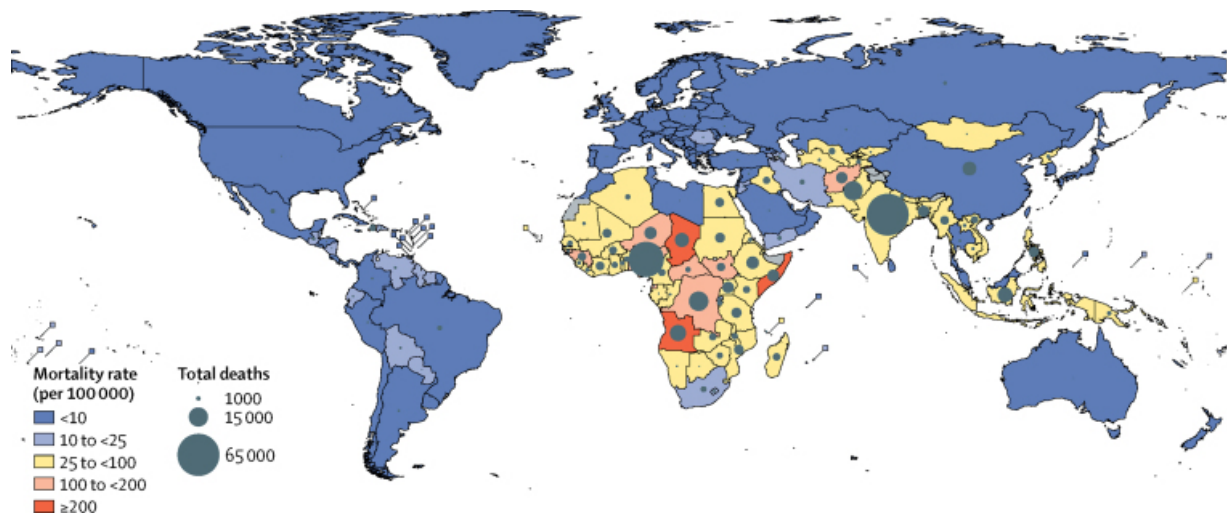
*Streptococcus pneumoniae* (the pneumococcus) is a Gram-positive bacterial pathogen that has played a historically important role in the fields of genetics and microbiology. The discovery of *S. pneumoniae* is attributed to George M. Sternberg in New Orleans, United States and J. Pasteur in Paris, France 1880 (Pasteur, 1881; Sternberg, 1881). Stenberg and Pasteur simultaneously and independently isolated the bacteria through animal passage experiments. In the subsequent years following its isolation, the pathogen was recognised as a causative agent of pneumonia, meningitis, and otitis media (Watson, Musher, Jacobson, & Verhoef, 1993). In 1928, Frederick Griffith demonstrated transformation in bacteria through a landmark experiment using a nonvirulent pneumococcal strain and a heat killed virulent pneumococcal strain in murine models (Griffith, 1928). The live nonvirulent pneumococcal strain was able to transform into a lethal form when injected with a heat killed virulent strain in mice. Up until about 15 years later, the transformation factor was widely speculated to be protein. In 1944, again using *S. pneumoniae*, Avery, Macleod, and McCarthy (Avery, Macleod, & McCarty, 1944) demonstrated in a pioneering study that DNA is the hereditary material in living organisms and the transforming factor in the earlier study of Griffith. This work by Avery and colleague also highlighted the pneumococcal polysaccharide capsule as an important virulence factor and is widely viewed as the birth of modern molecular biology. The pneumococcus was also important in early discoveries of serum therapy and humoral immunity. Albert Fraenkel discovered in 1886 that rabbits that survived pneumococcal infection were immune to reinfection (Fraenkel, 1886). This work was extended by the Klemperer brothers in 1891 showing that immunity conferred by heat-killed pneumococci was strain dependent and that infusion of serum from previously immunised rabbits protected against primary pneumococcal infection (Klemperer & Klemperer, 1891a, 1891b). Despite its early discovery, the mechanism by which *S. pneumoniae* caused disease is still not fully understood.

### **1.2 The epidemiology of *Streptococcus pneumoniae***

*S. pneumoniae* is a leading infectious cause of pneumonia, meningitis, and septicaemia worldwide (Henriques-Normark and Tuomanen 2013). The highest burden of pneumococcal disease is in children less than 5 years old with about half of all deaths



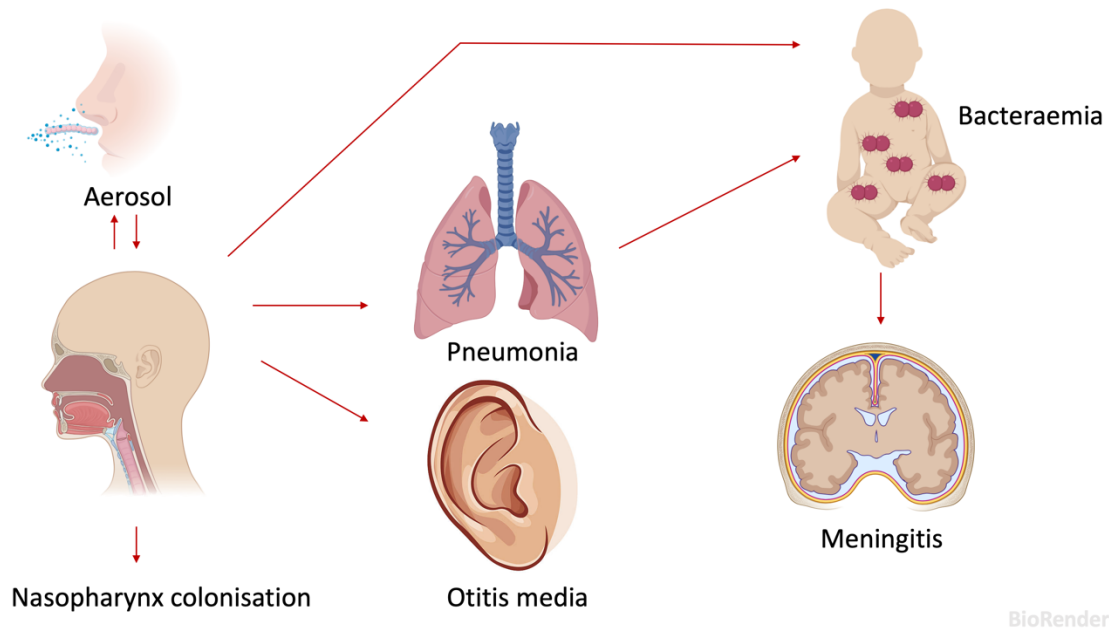
occurring in children living in sub-Saharan Africa (**Figure 1.1**) (Wahl *et al.* 2018). *S. pneumoniae* are diverse and at least 100 different capsular serotypes have been identified thus far (Henriques-Normark and Tuomanen 2013; Ganaie *et al.* 2020). Different pneumococcal serotypes have differing propensities to cause disease and thus vaccines are targeted against serotypes that most commonly cause invasive disease (Brueggemann *et al.* 2004; Ruckinger *et al.* 2009). The 13-valent pneumococcal conjugate vaccine (PCV-13) currently has the highest protection against 13 of the most common serotypes causing disease in children less than 5 years old (Feldman and Anderson 2020). Vaccines have been effective in reducing the burden of pneumococcal disease over many years (Feldman and Anderson 2020). However, high residual carriage of vaccine serotypes following PCV introduction have been observed in developing countries (Swarthout, Fronterre, *et al.* 2020; Usuf *et al.* 2021). Continuing vaccine effectiveness is also complicated by serotype replacement. Serotype replacement is a phenomenon whereby disease due to *S. pneumoniae* serotypes included in the vaccine is replaced by other serotypes not covered by the vaccines (Feldman and Anderson 2020; Love, Huber, and Anders 2014). Current vaccine strategies are limited by the number of serotypes that can be targeted. To compound this, disease caused by non-encapsulated pneumococci which cannot be targeted by current vaccines are emerging (Takeuchi *et al.* 2020; Keller, Robinson, and McDaniel 2016). As an alternative, capsule-independent strategies that target conserved proteins across pneumococcal strains or use of inactivated whole cells to eliminate the possibility of serotype replacement are being explored (Masomian *et al.* 2020). Due to the continued high burden of disease caused by *S. pneumoniae*, the WHO has designated *S. pneumoniae* a priority pathogen for research and development (WHO 2017).



**Figure 1.1:** Country-specific mortality rates and deaths estimates attributable to *S. pneumoniae* in HIV-negative children 1-59 months old in 2015 (Wahl *et al.* 2018).

### 1.3 *S. pneumoniae* host niches

The human upper respiratory tract (URT) is the natural niche of *S. pneumoniae* where it is carried by healthy individuals as a commensal (Bogaert, de Groot, and Hermans 2004). Nasopharyngeal colonisation by *S. pneumoniae* can first occur within a few days after birth (Hill *et al.* 2006). However, due to mechanisms that are not fully understood, *S. pneumoniae* can sometimes invade the URT epithelial barrier and translocate to the bloodstream to cause bacteraemia, to the meninges to cause meningitis, to the lungs to cause pneumonia, and to the middle ear to cause otitis media (**Figure 1.2**) (Bogaert, de Groot, and Hermans 2004; Henriques-Normark and Tuomanen 2013). Once in the bloodstream, pneumococcus can also disseminate to more diverse organs including the heart, kidneys, and spleen and cause organ dysfunction (Brown *et al.* 2014; Carreno *et al.* 2021; Huang *et al.* 2014; D'Mello *et al.* 2020). For example, pneumococcal invasion of cardiac tissue can lead to congestive heart failure, arrhythmias, and myocardial infarction which is triggered by microlesions in the myocardium (Brown *et al.* 2014; Subramanian, Henriques-Normark, and Normark 2019; Reyes *et al.* 2017).



**Figure 1.2:** Progression of pneumococcal disease showing its progression routes from being a commensal in the nasopharynx and translocating to different body parts to cause potentially fatal infections. Transmission of pneumococcus between individuals is believed to be mainly via aerosol droplets. Figure generated using Biorender (Perkel 2020).

#### 1.4 The nasopharyngeal environment and *S. pneumoniae*

*S. pneumoniae* colonisation of the nasopharynx is a prerequisite for the development of pneumococcal disease (Bogaert, de Groot, and Hermans 2004). The nasopharynx is a rich and diverse microenvironment which is resident to many other bacteria and viruses that compete with *S. pneumoniae* for available nutrients for survival (Man, de Steenhuijsen Piters, and Bogaert 2017). During colonisation of the nasopharynx, *S. pneumoniae* has to interact with and evade the host innate immune response (Weiser, Ferreira, and Paton 2018). Pneumococcal surface proteins for example block the activation of the host complement system and the ability of immunoglobulins to opsonise and clear the pathogen (Jedrzejewski 2001; Brooks and Mias 2018). In comparison to other host niches, the nasopharynx has several environmental and physiological differences in temperature, oxygen and carbon dioxide levels, major carbohydrate source, and concentration of transition metals (Man, de Steenhuijsen Piters, and Bogaert 2017; van Opijnen and Camilli 2012).

#### 1.4.1 Temperature

Temperature is one of the key physiological differences between the anatomical sites that *S. pneumoniae* occupies (Man, de Steenhuijsen Piters, and Bogaert 2017). The temperature of the nasopharyngeal epithelial surfaces, the natural niche of *S. pneumoniae*, is about 30°C (Lindemann *et al.* 2002). Following invasion of the epithelial surfaces, the pathogen encounters core body temperatures of around 37°C which can be as high as 40°C as part of the host febrile response to infection (Kwon *et al.* 2003; Aprianto *et al.* 2018).

#### 1.4.2 Carbohydrate source

*S. pneumoniae* depends on carbohydrate metabolism for energy and growth (Paixao, Caldas, *et al.* 2015) and a large part of its genome is dedicated to the uptake and metabolism of carbohydrates (Tettelin *et al.* 2001). Carbohydrates are essential for pneumococcal pathogenesis, and it is known to metabolise at least thirty-two different carbohydrates (Tettelin *et al.* 2001; Bidossi *et al.* 2012). *S. pneumoniae* also possess no fewer than ten extracellular glycosidases, which enable it to modify and breakdown a variety of hosts' glycans to generate free sugar (King 2010). The nasopharyngeal environment is poor in free sugars and the main carbon source in this environment is mucin on the mucosa lining (Paixao, Caldas, *et al.* 2015). N-acetylglucosamine (GlcNAc) makes up a third of mucin which is the main carbon source in the nasopharyngeal environment (Xia *et al.* 2005). Galactose is the second most abundant carbohydrate in the nasopharyngeal niche constituting about 29% of mucin (Xia *et al.* 2005). Other carbohydrates in mucin are fucose and sialic acid (Xia *et al.* 2005). In blood, the main carbohydrate source is glucose. When available, glucose has been shown to be the preferential substrate for pneumococcal metabolism (Paixao, Caldas, *et al.* 2015).

#### 1.4.3 Transition metals

Transition metals act as essential cofactors for many metabolic and enzymatic processes (Andreini *et al.* 2008). The acquisition of metal ions such as iron ( $\text{Fe}^{2+}$ ), manganese (Mn), and Zinc ( $\text{Zn}^{2+}$ ), is important for survival of *S. pneumoniae* (Weiser, Ferreira, and Paton 2018). Mn mediates *S. pneumoniae* resistance to oxidative stress (Honsa, Johnson, and

Rosch 2013) and is also involved in the ability of the pathogen to adhere to and colonise the nasopharynx (Rosch *et al.* 2009). During infection, the host chelates excess beneficial metals to inhibit availability for microbial growth (Corbin *et al.* 2008; Cassat and Skaar 2013; Kehl-Fie and Skaar 2010) and conversely releases toxic metals such as copper (Cu) and Zn to eliminate the pathogen (Honsa, Johnson, and Rosch 2013; Weiser, Ferreira, and Paton 2018; Cassat and Skaar 2013). This race for transition metals between the host and pathogen has been postulated to determine the success of bacterial pathogens (Honsa, Johnson, and Rosch 2013). The availability of transition metals varies between the different microenvironments that *S. pneumoniae* occupies (Honsa, Johnson, and Rosch 2013; McDevitt *et al.* 2011). The concentration of iron in the nasopharyngeal environment for example, is much lower than in blood, which is replete in iron mainly bound to haemoglobin (McDevitt *et al.* 2011). Iron is involved in the regulation of diverse cellular processes in *S. pneumoniae* such as amino acid synthesis, electron transport and cellular respiration (Honsa, Johnson, and Rosch 2013; Sheldon and Heinrichs 2015; Cassat and Skaar 2013) and iron overload or iron supplementation is associated with enhanced susceptibility to infection by bacterial pathogens (Cassat and Skaar 2013; Cross *et al.* 2015). In iron restricted environments, *S. pneumoniae* has been shown to increase protein expression of virulence factors such as pneumococcal surface antigen A (*psaA*) as part of its adaptive response (Nanduri *et al.* 2008). *S. pneumoniae* Iron uptake genes *piuA* and *piaA* are key bacterial factors that are required for pneumococcal virulence and colonisation (Brown, Gilliland, and Holden 2001; Weiser, Ferreira, and Paton 2018).

#### **1.4.4 Other physiological properties of the nasopharyngeal niche**

The nasopharyngeal conditions such as lower temperature, limited nutrient availability, and interaction with epithelial cells encourage biofilm formation and increase transformation efficiency making the nasopharynx the primary site for exchange of genetic material and fitness selection between *Streptococci* (Marks, Reddinger, and Hakansson 2012). This is mediated through permanent competence induction alongside downregulation of capsular genes (Marks, Reddinger, and Hakansson 2012). As the conditions of the nasopharyngeal environment promote *S. pneumoniae* survival and fitness, the potential dead-end associated with translocation to the meninges or blood

and probable death of the host may be viewed as accidental for the pathogen (Hava, LeMieux, and Camilli 2003).

### 1.5 Adaptation and virulence factors

Adaptation of bacterial pathogens to new environments requires coordinated regulation of gene expression (Hava, LeMieux, and Camilli 2003; Colgan, Cameron, and Kroger 2017). A major knowledge gap in pneumococcal biology is understanding of the mechanisms by which the pneumococcus is able to transform from a commensal in the upper respiratory tract to a pathogen, translocating to the lungs, brain, and bloodstream where it often results in serious infections which may be fatal. Knowledge of the genetic factors that make certain serotypes to be more invasive and more likely to cause disease than other serotypes also remain incomplete. Although the host immune system is an important factor in determining outcomes of pneumococcal colonization and disease (Bogaert, de Groot, and Hermans 2004; Igartua *et al.* 2017; Brooks and Mias 2018), changes in the pathogen's gene expression, metabolic lifestyle, and genetic makeup are the main determinants (Lees *et al.* 2017). The capsule type of strains (serotype) plays an important and stable role in determining pneumococcal disease outcomes (Weinberger *et al.* 2010; Brueggemann *et al.* 2004; Balsells *et al.* 2018). Host immune evasion and subversion by *S. pneumoniae* is mediated by expression of several bacterial virulence factors such as the capsular polysaccharide (*cps*) which is a major pneumococcal virulence factor (Weiser, Ferreira, and Paton 2018). Pneumococcal virulence factors and their functions have been identified predominantly through single gene mutant knockout and murine infection studies (Canvin *et al.* 1995; Watson, Musher, and Verhoef 1995; Manco *et al.* 2006; Pracht *et al.* 2005; Ogunniyi *et al.* 2007; Jedrzejewski 2001). Mutant knockout studies, which assume highly controlled and specific deletion of bacterial genes, are usually undertaken to identify phenotypes linked to the absence of genes. Phenotypic difference of the knockout mutants' strains compared to wild type strains, usually in murine models, is used to infer function of knocked out genes. **Table 1.1** Is a selected list of pneumococcal virulence factors that have been identified to contribute to pneumococcal virulence.

**Table 1.1:** A selection of major *S. pneumoniae* virulence factors

Virulence factor	Function	Ref
<i>cps</i>	Prevent mucus entrapment and Inhibit opsonophagocytosis	(Weiser, Ferreira, and Paton 2018)
<i>ply</i>	Proinflammatory, cytotoxic, and pro-apoptotic toxin	(Weiser, Ferreira, and Paton 2018; Brooks and Mias 2018; Zafar <i>et al.</i> 2017)
<i>psaA</i>	Manganese uptake Resistance to oxidative stress	(Brooks and Mias 2018)
<i>cbpA</i>	Adherence and invasion	(Brooks and Mias 2018)
<i>lytA</i>	Autolysin that digests cell wall	(Brooks and Mias 2018)
<i>nanA and nanB</i>	Enhanced colonisation	(Brooks and Mias 2018)
<i>pilA, pilA, and pilA</i>	Iron acquisition	(Brown, Gilliland, and Holden 2001; Brown <i>et al.</i> 2001; Whalan <i>et al.</i> 2005, 2006)
<i>bgaA</i>	Adherence and carbohydrate degradation	(Weiser, Ferreira, and Paton 2018)
<i>sodA</i>	Resistance to oxidative stress	(Weiser, Ferreira, and Paton 2018)
<i>pavA</i>	Adherence and invasion	(Weiser, Ferreira, and Paton 2018)

## 1.6 Microarrays

Gene expression analyses have enhanced our understanding of pathogenic mechanisms, pathogen responses to environmental stimuli, and host-pathogen interactions (Miller and Tang 2009). Studies of the global gene expression changes in bacteria were first made possible with the advent of microarrays in 1999, providing unprecedented insight into transcriptional response of *Escherichia coli* (Tao *et al.* 1999; Richmond *et al.* 1999). Microarrays exploit the complementary base pairing property of nucleic acids. A microarray chip contains several fluorescently labelled nucleic acid probes (or oligos) immobilised as discrete features or spots to a solid surface (Miller and Tang 2009). “Unknown” samples containing sequences of interest are hybridized after washing steps and the amount of binding of samples to the probes is quantified by the intensity of the fluorescent signal for a particular spot using a fluorescent scanner (Miller and Tang 2009). Microarrays studies have expanded our knowledge of gene expression in both research settings and clinical diagnostics, leading to an exponential growth in number of publications using the technology that have infectious disease or microbiology applications (Miller and Tang 2009). Microarray technology has been applied to microbial pathogens detection and identification (You *et al.* 2008), antimicrobial resistance monitoring (Call *et al.* 2003), and strain typing (Satzke *et al.* 2015). However, one of the major disadvantages of microarrays methods for gene expression analyses is the requirement for a known and well annotated genome sequence to make a pre-established oligonucleotide array. Furthermore, because the technology is based on hybridization, the dynamic range and specificity of detection is limited.

## 1.7 RNA sequencing (RNA-Seq)

RNA sequencing (RNA-Seq) addresses the limitations of microarrays and has transformed investigations into gene expression of bacterial pathogens by enhancing our understanding of bacterial transcriptomic features (Kroger *et al.* 2013; Aprianto *et al.* 2018; Kroger *et al.* 2018; Nicolas *et al.* 2012; Filannino *et al.* 2018; Shah 2014; Caglar *et al.* 2017; Colgan, Cameron, and Kroger 2017). Advantages of RNA-Seq over hybridization methods such as microarrays for gene expression studies include improved sensitivity and higher dynamic range since transcription is studied at a single nucleotide level (van Vliet



2010). As well as being capable of annotation-independent detection of transcription (Croucher and Thomson 2010; Colgan, Cameron, and Kroger 2017), RNA-seq presents a methodology to identify previously unknown transcriptomic features such as small noncoding RNA (sRNA) species and can help to improve annotation of genome sequences through the identification of promoter regions and 5' and 3' boundaries of mRNA species (van Vliet 2010).

### 1.8 RNA-Seq for bacterial transcriptomics studies

Using the model microorganism *Salmonella enterica* serovar Typhimurium, Kroger *et al* (Kroger *et al.* 2013) demonstrated the applicability of RNA-Seq in deciphering the transcriptomic regulation of bacterial pathogens in tightly controlled *in vitro* conditions that the authors termed “infection-relevant conditions”. These conditions comprise of various physiological and biochemical stressors which reflect aspects of the infection process of the *Salmonella*. The study highlighted the environmental conditions that induce key virulence factors such as *Salmonella* pathogenicity Island genes, and sRNA expression of the pathogen. This study was one of the early studies to demonstrate the utility of RNA-Seq in deciphering the gene expression patterns of bacterial pathogens and providing new insights into the biology of *Salmonella* (Kroger *et al.* 2013). A similar methodology was applied to the important emerging bacterial pathogen *Acinetobacter baumani*, characterising for the first time the transcriptomic architecture and sRNA repertoire of the pathogen (Kroger *et al.* 2018). Combining RNA-seq profiles from 108 experiments of two clinical *Staphylococcus aureus* strains, Poudel *et al* (Poudel *et al.* 2020) showed a model of independently modulated sets of genes linked to various cellular processes that underlie physiological responses during host infection. Comparative transcriptomics using RNA-Seq for two genetically similar *Salmonella* strains has also enabled the characterization of the differential gene expression of virulence and metabolic genes under the same conditions and led to identification of a novel plasmid, a cysteinyl-tRNA synthetase (*cysS*) (Canals *et al.* 2019).

### 1.9 Using RNA-Seq to study *S. pneumoniae* transcriptomics adaptation

The first study to investigate *S. pneumoniae* gene expression similar to Kroger *et al* (Kroger *et al.* 2013; Kroger *et al.* 2018) was conducted by Aprianto *et al* (Aprianto *et al.* 2018) where they quantified the transcriptomic response of *S. pneumoniae* D39 strain (serotype 2) to 22 different conditions. This work showed a high level of dynamic expression with all annotated genomic features of *S. pneumoniae* being expressed in at least one of the studied conditions (Aprianto *et al.* 2018). In addition, they were able to identify a previously unknown member of pneumococcal competence regulon (Aprianto *et al.* 2018). Expanding on this work, Sinha *et al* (Sinha *et al.* 2019) and Slager *et al* (Slager, Aprianto, and Veening 2019) identified multiple novel sRNAs of *S. pneumoniae* using published bacterial RNA-Seq datasets. Although the original study (Aprianto *et al.* 2018) provided an important roadmap for pneumococcal gene transcription, the authors relied on the D39 (serotype 2) laboratory strain descended from the original Avery experiment many decades ago (Avery, Macleod, and McCarty 1944). The D39 strain causes disease in murine infection models was first isolated in 1914 and has been passage ever since under laboratory conditions (Griffith 1928; Slager, Aprianto, and Veening 2018). The D39 strain behaves differently from clinical human *S. pneumoniae* strains under the same conditions (Pozzi *et al.* 1996). It is possible that the evolutionary physiological and metabolic properties of D39 are not reflective of currently circulating clinical *S. pneumoniae* strains causing disease in humans. *S. pneumoniae* exhibits considerable plasticity and pneumococci are known for their capacity to remodel their genomes through the uptake and incorporation of exogenous DNA from other closely related pathogens (Weiser, Ferreira, and Paton 2018). Up to a tenth of genes have been found to be divergent or absent when the historical laboratory adapted strain R6, a serotype 2 derivative of D39 strain, was compared to a more recent clinical serotype 4 strain (TIGR4) (Bruckner *et al.* 2004). In fact, comparison of two D39 (serotype 2) strains, D39(Lilly) and D39(NCTC), showed major mutations that affected key metabolic and regulatory genes which resulted in differing virulence of the strains (Lanie *et al.* 2007). In addition, 15% of putative small RNAs (sRNAs) in a clinical serotype 4 strain (TIGR4) were absent in the D39 (serotype 2) laboratory strain (Sinha *et al.* 2019). Some of these pneumococcal sRNAs have been previously implicated in TIGR4 virulence and suggested to contribute to the invasive potential of TIGR4 clinical strains (Sinha *et al.* 2019). How the transcriptomic profile of *S.*

*pneumoniae* changes over time for some conditions was not described in the D39 transcriptomic study (Aprianto *et al.* 2018). Thus, quantifying *S. pneumoniae* gene expression at different timepoints may provide important insights on how the bacterium fine-tunes and evolves its gene regulation in response to environmental stresses. Using mutagenesis approaches, a particularly strong association has been described for a three-gene pneumococcal operon between *in vitro* metal stress conditions (FeSO<sub>4</sub> and MnSO<sub>4</sub>) and colonisation of murine nasopharynx (van Opijnen and Camilli 2012). This three-gene operon, which was shown to be essential for colonisation, but not for lung infection, was later identified as the *cop* operon (*copAY*, *cupA*) (van Opijnen and Camilli 2012; Shafeeq *et al.* 2011). The *cop* operon is a virulence factor for pneumococcal colonisation and is involved in regulating pneumococcal exposure to elevated copper levels (Shafeeq *et al.* 2011). However, the use of RNA-Seq to investigate the genome-wide transcriptomics response of *S. pneumoniae* to transition metals exposure has not been attempted. Transition metal ions are critical for several bacterial biological processes and have different concentrations in the host niches *S. pneumoniae* occupies (McDevitt *et al.* 2011; Honsa, Johnson, and Rosch 2013; van Beek *et al.* 2020). Transcriptomic response of pneumococcus to metal stress will provide valuable insights into how the bacteria adapts to similar host environments. Furthermore, there are currently no published studies comparing transcriptomic response of different *S. pneumoniae* serotypes. This thesis aims to extend the work by Aprianto *et al.* (Aprianto *et al.* 2018) by comparing the transcriptomic response of *S. pneumoniae* strains with different propensities to colonise or cause invasive disease in controlled biologically relevant environments. Therefore, this work may help identify key candidate genes or biological pathways that predispose some *S. pneumoniae* serotypes to a more pathogenic lifestyle.

### 1.10 Thesis aims

Understanding the molecular mechanisms behind the pathogenesis of different *S. pneumoniae* serotypes during colonisation and invasive disease is crucial to developing novel therapeutics. Colonisation of the URT is a prerequisite for the development of pneumococcal disease. However, *S. pneumoniae* serotypes differ in their propensity to be carried in the host URT and to cause invasive disease. Differences in adaptation of *S. pneumoniae* strains to the environment conditions of the URT may contribute to these differences. We hypothesize that *S. pneumoniae* has a characteristic adaptive transcriptomic profile to a range of biologically relevant environmental conditions that vary by propensity of strains to colonise or cause invasive disease in the host. The overall aim in this thesis is therefore to evaluate the genome-wide transcriptomic response of clinical pneumococcal strains to variations such as temperature, metal ions concentration, and carbohydrate source *in vitro* and then investigate these transcriptomic profiles *in vivo*.

The specific aims were:

- Identify an optimal chemically defined medium for study of the transcriptomic response of clinical *S. pneumoniae* strains to different environmental exposures.
- Optimise and establish experimental and analysis methods to conduct comparative RNA-Seq transcriptomics analysis of multiple clinical strains of *S. pneumoniae*.
- Characterise the gene expression profiles and regulation under environmental conditions associated with niche adaptation and evaluate differences between strains varying in clinical phenotype (serotype 1 and serotype 6B).
- Generate high-quality genome assemblies of clinical pneumococcal strains for mapping and quantification of RNA-Seq sequence reads.

- Explore the application of *in vitro* methods developed in this thesis for investigating the transcriptome of *S. pneumoniae* retrieved from human upper respiratory tract.

## Chapter 2 : General materials and methods

## 2.1 Chemicals and reagents

Unless otherwise stated, all reagents and chemicals used were reagent grade and purchased from Sigma-Aldrich (Merck, Germany).

## 2.2 General safety and control measures

All bacterial culture work was undertaken in biosafety cabinet class II. The required personal protective equipment's (PPE) were worn for all laboratory work and all safety protocols for working with potentially hazardous agents were followed in line with University College London (UCL) guidelines.

## 2.3 *S. pneumoniae* strains and rationale for selection

The *S. pneumoniae* strains used for growth curve analysis in Chapter 3 are summarised in **Table 2.1**. The clinical strain BHN418 (serotype 6B) (Browall *et al.* 2014) was selected because: (a) It is a commonly carried serotype that also causes life-threatening diseases; (b) it is the strain used in the experimental human pneumococcal carriage (EHPC) model at Liverpool School of Tropical Medicine (LSTM) (Gritzfeld *et al.* 2013), which we plan to use to investigate the pneumococcal transcriptome during colonisation *in vivo* (Chapter 8); (c) it is the parent strain for a live attenuated pneumococcal vaccine for which clinical trials are being carried out at LSTM (Chapter 5). TIGR4 (serotype 4) is the first clinical *S. pneumoniae* strain that was whole genome sequenced by Tettelin and colleagues in 2001 (Tettelin *et al.* 2001). *S. pneumoniae* strains P1262 (serotypes 15B) and P1121 (serotype 23F) were also included because they are common serotypes and were used as challenge strains in the EHPC model. Serotype 15B has been emerging as common serotype replacement strain (Lo *et al.* 2019). D39 (serotype 2) is a laboratory *S. pneumoniae* strain multiply passaged that is commonly used in pneumococcal *in vitro* and murine *in vivo* experiments and was investigated in a large RNA-Seq screen (Aprianto *et al.* 2018). Strain BVJ11JL (serotype 1) is a typically virulent pneumococcal strain. Serotype 1 has been responsible for outbreaks and epidemics in sub-Saharan Africa (Balsells *et al.* 2018). Serotype 1 is one of the most commonly isolated serotypes in invasive pneumococcal

disease but unlike serotype 6B for example, it is rarely isolated in the nasopharynx of healthy individuals (Heffron 1939; Brueggemann *et al.* 2004).

**Table 2.1:** Summary description of *S. pneumoniae* strains used in growth curve analysis

Strain	Serotype	PCV13 serotype	Obtained from	Origin (Reference)
BHN418	6B	Yes	Jeremy S Brown	Sweden (Browall <i>et al.</i> 2014)
BVJ1JL	1	Yes	Todd Swartwout	Malawi (Swarthout, Fronterre, <i>et al.</i> 2020)
D39	2	No	Jeremy S Brown	USA (Avery, Macleod, and McCarty 1944)
P1262	15B	No	LSTM	NA
P1121	23F	Yes	LSTM	USA (McCool <i>et al.</i> 2002)
TIGR4 (P1672)	4	Yes	Jeremy S Brown	Norway (Hyams <i>et al.</i> 2010)

## 2.4 Bacterial stocks preparation

Apart from the BVJ1JL strain (serotype 1), all other strains were obtained as purified *S. pneumoniae* single colonies from the laboratories listed in **Table 2.1**. To generate in-house stocks for the work in this thesis, single colonies of purified strains were plated on Columbia blood agar (CBA) (Oxoid) supplemented with 5% horse blood and incubated at 37°C and 5% CO<sub>2</sub>. Cultures were confirmed to be alpha-haemolytic *Streptococci* by morphological examination of cultures following ~18hrs incubation. *S. pneumoniae* bacterial stocks were prepared by growing cultures in Todd Hewitt broth (Oxoid) + Yeast (THY) media to early-log phase (OD<sub>600</sub> 0.3) before adding 10% glycerol and stored in 1mL aliquots at -80°C. The number of colonies forming units (CFU) was calculated for each bacterial stock preparation at time of use. Aliquots of stocks were defrosted and used once.



### 2.5 Isolation and identification of *S. pneumoniae* serotype 1 (strain BVJ1JL)

The BVJ1JL strain was isolated in 2015 from a nasopharyngeal swab (NPS) obtained from a healthy 9-year-old Malawian child during community carriage surveillance in Blantyre, Malawi (Swarthout, Fronterre, *et al.* 2020). Being age-ineligible for the 13-valent pneumococcal conjugate vaccine (PCV13) infant immunisation introduced in 2011, the child was not PCV-vaccinated. The nasopharyngeal swab (NPS) was stored in skim milk-tryptone-glucose-glycerol (STGG) transport medium frozen at -80°C. The BVJ1JL NPS sample was selected because it was previously found by DNA microarray, whole-genome sequencing, and latex agglutination methods to be a single carriage serotype 1 sample (Swarthout, Gori, *et al.* 2020). The NPS of BVJ1JL was first grown in Columbia blood agar (Oxoid) (CBA) supplemented with 5% horse blood and incubated at 37°C with 5% CO<sub>2</sub> for ~18hrs. A single colony morphologically resembling alpha haemolytic *Streptococci* was picked and purified by plating in a fresh CBA plate and incubated at 37°C with 5% CO<sub>2</sub> for ~18hrs. The BVJ1JL strain was then confirmed to be *Streptococcus pneumoniae* by optochin test (sensitive), bile solubility test (Soluble), and Gram stain (Gram positive) (Werno and Murdoch 2008). The serotype of the BVJ1JL was confirmed by serological latex agglutination test kit (ImmuLex™ 7-10-13-valent Pneumotest; Statens Serum Institute, Denmark) and later by whole genome sequencing using PneumoCat (Kapatai *et al.* 2016). Confirmed single colonies of BVJ1JL were grown to OD<sub>600</sub> 0.3 in THY and 1mL stocks prepared and stored as described above (Section 2.4).

#### 2.5.1 Colony forming unit of cultures

Colony forming units (CFU/mL) of *S. pneumoniae* bacterial cultures were enumerated by serially diluting original liquid cultures with sterile phosphate-buffered saline (PBS) at concentrations ranging from of 10<sup>-1</sup> to 10<sup>-6</sup> in final volume of 200μL. 20μL of each dilution factor was placed into CBA plate supplemented with 5% horse blood using pipettes in triplicates. The culture was allowed to dry on the CBA plate at room temperature and then incubated at 37°C with 5% CO<sub>2</sub> for ~18hrs. The CFU/mL of the original culture was calculated by using the average number of visible colonies from replicates at a specific dilution factor using the formula.

CFU/mL = (Average number of colonies x dilution factor) / volume added on culture plate

## 2.6 Growth media

Selection of a chemically defined medium (CDM) that can support both pneumococcal growth and enable alterations in the constituents of the medium in a controlled manner was a crucial step for the planned transcriptomics experiments in this thesis. Five different growth media were evaluated.

### 2.6.1 Todd Hewitt broth + Yeast (enriched media)

Todd Hewitt broth + Yeast (THY) is a common enriched media used for growth of *S. pneumoniae* (Todd and Hewitt 1932; Sanchez-Rosario and Johnson 2021). Todd Hewitt broth (Oxoid) and yeast extract (Oxoid) were purchased separately in a powder form. To make THY media, 18.4g Todd Hewitt broth (Oxoid) and 2.5g yeast extract (Oxoid) were added to 500mL distilled water and autoclaved at 121°C. The prepared THY medium was stored at room temperature with maximum shelf life of 2 weeks. THY was used as control medium for growth comparison to four CDMs and was also evaluated as an iron replete growth medium in experiments evaluating effect of iron limitation on *S. pneumoniae* growth and transcriptome (Chapter 6).

### 2.6.2 7H9+ medium

The 7H9+ medium was developed by Dr Lavanya Mane, a previous PhD student in the Heyderman laboratory at UCL. Based on a standard mycobacterium growth media, 7H9+ medium was optimised by Dr Mane to enable growth of both *S. pneumoniae* strain TIGR4 (serotype 4) and *Mycobacterium tuberculosis* for metabolic profiling. The 7H9+ medium was made using the commercially available Middlebrook 7H9+ broth base (Sigma-Aldrich) supplemented with vitamins and amino acids from Sicard's medium recipe (Sicard 1964). The detailed protocol for preparation of 7H9+ medium is presented in **Appendix 2**.

### 2.6.3 Nose mimicking conditions (NMC) medium

NMC is a chemically defined media described by Aprianto *et al* (Aprianto *et al.* 2018) and used to investigate transcriptomic response of *S. pneumoniae* D39 strain (serotype 2) in

various environmental condition. NMC is an adapted modification of Sicard's defined medium (Sicard 1964). The detailed protocol for preparation of NMC as described by (Aprianto *et al.* 2018) is reproduced in **Appendix 3**.

#### 2.6.4 Lung mimicking conditions (LMC) medium

LMC was also described by Aprianto *et al* (Aprianto *et al.* 2018) and is similar to NMC except for concentration of glucosamine (1.28g/l for NMC vs 0.64g/l for LMC) and serum albumin (1g/l for NMC and 3g/l for LMC) (**Appendix 3**).

#### 2.6.5 Chang CDM

Described by Chang *et al*, 2011 (Chang, LaSarre, *et al.* 2011) is an improved and streamline chemically defined media for the growth of *Streptococci* that was adapted and modified from van de Rijn *et al* (van de Rijn and Kessler 1980). The CDM was prepared as described in supplementary protocol S1 in the published article (Chang, LaSarre, *et al.* 2011).

### 2.7 Growth curve assays

Frozen aliquots of *S. pneumoniae* bacteria stocks ( $\sim 1 \times 10^8$  cells) were thawed from  $-80^\circ\text{C}$ . The bacterial cells were initially washed twice with equal volume of PBS by centrifugation (8000g, 10 min,  $4^\circ\text{C}$ ) to remove excess THY and glycerol residues (subsequently changed – see results Chapter 3). In initial experiments, 300 $\mu\text{L}$  of bacterial suspension was added to 25mL of media in 50mL falcon tubes. The tube screw caps were loosely tightened and incubated at  $37^\circ\text{C}$  and 5%  $\text{CO}_2$ . Optical density ( $\text{OD}_{600}$ ), used as a proxy for bacterial growth, and measured every 1hr in 1mL clear curvets using a benchtop spectrometer (Thermo Scientific).

Growth curve data was also obtained using the automated Spark<sup>®</sup> multimode plate reader (Tecan). The plate reader enabled measuring of multiple timepoints in a more controlled temperature environment over a longer duration and reduced the laborious nature of measuring growth in 50mL falcon tubes. For growth curves with plate reader, 300 $\mu\text{L}$  of *S. pneumoniae* bacterial stock ( $\sim 1 \times 10^8$  cells) was added to 25mL media in 50mL falcon tubes and then 200 $\mu\text{L}$  of cultures were aliquoted in a sterile U-bottom culture plate (Cellstar)

with a magnetic lid attached. The plate reader was programmed to measure OD<sub>600</sub> every 30 minutes following 5 seconds shaking over an 18hr period. For each condition, at least three (and up to nine) technical replicates were included.

## **2.8 Growth conditions**

Growth variations in three different experimental conditions were tested: temperature, carbohydrate source, and transition metals concentration. The different concentrations used for each condition are described in more detail in each Chapter 3. Unless otherwise stated, three independent experiments were conducted for each growth condition. At least 3 technical replicates were included for each experiment

## **2.9 Growth parameters analysis**

Data from plate reader was exported into Microsoft Excel 2016 to calculate mean growth OD for each timepoint and standard error from the mean (SEM) across replicates. This data was then imported into R software (Version 3.4.2) to generate graphs and conduct statistical analysis. Kruskal-Wallis one-way analysis was done to evaluate statistical differences between groups in growth curve experiments. The Curveball algorithm (Ram *et al.* 2019) was used to calculate growth parameters for the various growth conditions tested. The differences in growth parameters between conditions were statistically evaluated using Kruskal-Wallis non-parametric analysis.

## **2.10 RNA isolation protocol**

This RNA isolation protocol employs chemical and mechanical techniques to lyse *S. pneumoniae* cells using the commercially available MirVana miRNA extraction kit (Invitrogen). The protocol was adapted and modified from Mann *et al* (2012) (Mann *et al.* 2012). The protocol detailed below is the optimised version for RNA isolation of *S. pneumoniae* from *in vitro* experiments used in Chapters 5-8. Good laboratory practice (GLP) guidelines for isolation of RNA were followed throughout.

### **Equipment and reagents for RNA isolation**

- Precellys tissue homogenizer (Bertin)

- Precellys VK01 50 preps bead beating tubes
- Heat block / water bath
- RNA isolation hood/ Safety cabinet class II
- MirVana miRNA extraction kit (Invitrogen)
- DNase digestion kit (TURBO DNA-free Kit-50 reaction (Invitrogen))
- Agilent Bioanalyzer RNA 6000 Nano reagents and chips
- Mutanolysin (Sigma)
- Lysozyme powder (Sigma)
- Qiashredder (Qiagen)
- QIAGEN RNeasy MinElute Cleanup Kit (Qiagen).
- 100% ethanol (Sigma)

### **RNA isolation set up**

1. General RNA-handling considerations were observed throughout.
  - All areas/kit RNase-zapped regularly.
  - All plasticware used were RNase free.
2. Collect ice in polystyrene box.
3. Collect samples from -80°C freezer location and place in hood.
4. Set water bath / heat block to 70°C.
5. Set samples to thaw in Ice.
6. Clean hood and work area and equipment's with 70% alcohol and RNA-zap regularly.
7. Label tubes for Precellys, Qiashredder and Mirvana extraction and keep on bench.

### **MirVana RNA isolation preparation**

8. Add 21mL of 100% ethanol to the bottle labelled miRNA Wash Solution 1. Mix well.  
Place a check mark in the empty box on the label to indicate that ethanol has been added. Store at 4°C and equilibrate to room temperature before use.
9. Add 40mL of 100% ethanol to the bottle labelled Wash Solution 2/3. Mix well.  
Place a check mark in the empty box on the label to indicate that the ethanol has been added.

10. Place Lysis/Binding Buffer on ice to keep at 4°C.
11. Place miRNA Homogenate Additive on ice to keep at 4°C.
12. Place Acid-Phenol: Chloroform mixture on ice to keep at 4°C.
13. Place 1.4 mL Gel Loading Buffer II from -20°C on ice to thaw at 4°C.
14. Get 5mL Elution Solution at room temperature ready.
15. Place DNase kit on ice to thaw.
16. Prepare Lysis solution containing 50U mutanolysin and 30mg/mL of lysozyme in lysis/binding buffer. 1mL lysis solution required for each sample. For 20 samples for example, add 200µL of 5KU mutanolysin and 900mg lysozyme to 19.8mL of lysis/binding buffer. Vortex to dissolve lysozyme in solution.

### **Sample preparation**

1. Thaw samples on bench if frozen in RNA protection media. Go to step 5 if pellets frozen.
2. Spin sample in centrifuge for 10 minutes at 10,000-13,000 x g at 4°C.
3. Remove the RNA protection media from the pellets.
4. Add 1mL of lysis solution to pellets. See lysis solution preparation above.
5. Mix using pipetting to ensure all the pellet is dissolved.
6. Place the lysed sample in Precellys bead beating tubes.
7. Put all tubes on ice while preparing Precelley's bead beating machine.

### **Mechanical cell wall disruption and capsule debris removal**

1. Set Precellys machine speed to 6200 rpm for 45 seconds.
2. Put samples in machine, set top securely on top of sample.
3. Close lid and beat x3 at 20 seconds intervals.
4. Place homogenate in tube in water bath/ heat block set at 70°C for 10 minutes.
5. Put samples on ice to cool pre-spin.
6. Spin at 12,000 x g, 5 min, 4°C in pre-cooled microfuge.
7. Remove supernatant (lysate), avoid removing beads.
8. Place Up to 700µL of lysate in labelled QIAshredder spin column placed in a labelled 2mL RNase free collection tube.

9. Spin for 2 minutes at maximum speed in a pre-cooled microcentrifuge at 4°C. The lysate is homogenized as it passes through the spin column.

### **RNA extraction**

1. Increase temperature of water bath/heat block to 95°C for elution.
2. Add 1/10 volume of miRNA Homogenate Additive to the cell or tissue lysate (or homogenate) and mix well by vortexing or inverting the tube several times. For example, if the lysate volume is 700µL, add 70µL miRNA Homogenate Additive.
3. Leave the mixture on ice for 10 min. Change the temperature of the centrifuge to RT.
4. Add a volume of Acid-Phenol: Chloroform that is equal to the lysate volume before addition of the miRNA Homogenate Additive. For example, if the original lysate volume was 700µL, add 700µL Acid-Phenol: Chloroform.

#### **IMPORTANT!**

Be sure to withdraw from the bottom phase in the bottle of Acid-Phenol: Chloroform because the upper phase consists of an aqueous buffer.

5. Vortex for 30 - 60 seconds to mix thoroughly.
6. Centrifuge for 5 minutes at maximum speed (10,000 x g) at room temperature to separate the aqueous and organic phases. After centrifugation, the interphase should be compact; if it is not, repeat the centrifugation.
7. Carefully remove the aqueous (upper) phase without disturbing the lower phase and transfer it to an Eppendorf 1.5mL tube. Note volume removed.
8. Preheat elution solution or nuclease-free water to 95°C in the pre-heated water bath for use in eluting.
9. Add 1.25 volumes of room temperature 100% ethanol to the aqueous phase recovered in step 7 (e.g., if 640µL was recovered in step 7, add 800µL ethanol). Mix thoroughly with pipetting.
10. For each sample, place a filter cartridge into one of the collection tubes supplied.
11. Pipet the lysate/ethanol mixture (from the previous step) onto the filter cartridge. Up to 700µL can be applied to a filter cartridge at a time, for samples larger than this, apply the mixture in successive applications to the same filter.

12. Centrifuge for ~15 sec to pass the mixture through the filter centrifuge at RCF 10,000 x g.
13. Discard the flow-through and repeat until all the lysate/ethanol mixture is through the filter. Reuse the collection tube for the washing steps.
14. Apply 700µL miRNA Wash Solution 1 (working solution mixed with ethanol) to the filter cartridge and centrifuge for ~5–10 sec.
15. Discard the flow-through from the collection tube and replace the filter cartridge into the same collection tube.
16. Apply 500µL Wash Solution 2/3 (working solution mixed with ethanol) and draw it through the filter cartridge as in the previous step.
17. Repeat with a second 500µL aliquot of Wash Solution 2/3.
18. After discarding the flow-through from the last wash, replace the filter cartridge in the same collection tube and spin the assembly for 1 min at maximum speed to remove residual fluid from the filter.
19. Transfer the filter cartridge into a fresh collection tube (provided with the kit). Apply 100µL of pre-heated (95°C) Elution Solution or nuclease-free water to the centre of the filter and close the cap. Spin for ~20–30 sec at maximum speed to recover the RNA.
20. Collect the eluate (which contains the RNA) and store it at –20°C, or place on ice for DNase digestion.

### **DNA digestion**

1. To each ~100µL extracted sample, add:
  - 10µL TURBO DNase buffer (0.1 vol).
  - 1µL TURBO DNase enzyme.
2. Flick mix and briefly spin down in microfuge.
3. Incubate at 37°C in heat block for 30 mins.
4. Vortex DNase Inactivation Reagent and add 0.1 volumes per sample.
5. Incubate at RT for 5 min with occasional flick mixing.
6. Spin for 1.5 minutes at 10,000 x g.
7. Remove supernatant and put in a fresh microcentrifuge tube. Digested DNA will be in an invisible pellet.



8. Run sample through Qiagen RNeasy mini elute clean-up procedure before Bioanalyzer quantification.

### **RNeasy MinElute Cleanup Kit**

1. Add RNase-free water to RNA to bring the volume to 100 $\mu$ L (if not eluted to 100 $\mu$ L).
2. Add 350 $\mu$ L Buffer RLT. Pipette to mix.
3. Add 550 $\mu$ L 96%–100% ethanol. Pipette to mix.
4. Add half the sample mixture (~500 $\mu$ L) to a RNeasy MinElute spin column placed in a 2mL collection tube.
5. Close the lid, and then centrifuge at 8,000  $\times$  g for 15 seconds.
6. Discard the flow-through.
7. Add the remaining sample to the spin column.
8. Repeat steps 4–6.
9. Discard the flow-through and collection tube.
10. Place the RNeasy MinElute spin column in a new 2mL collection tube.
11. Add 500 $\mu$ L Buffer RPE, and then close the lid.
12. Centrifuge at 8000  $\times$  g for 15 seconds.
13. Discard the flow-through and retain the collection tube.
14. Add 500 $\mu$ L 80% ethanol to the RNeasy MinElute spin column, and then close the lid.
15. Centrifuge at 8,000  $\times$  g for 2 minutes.
16. Without allowing the column to contact the flow-through, remove the RNeasy MinElute spin column from the collection tube. Contact between the column and flow-through causes ethanol carryover.
17. Discard the flow-through and collection tube.
18. Place the RNeasy MinElute spin column in a new 2mL collection tube.
19. Place the spin columns into the centrifuge with at least 1 empty position between columns, which avoids damaging the spin column lids. Orient the lids to point in a direction opposite the rotation of the rotor. For example, if the rotor rotates clockwise, orient the lids counterclockwise.

20. With the lids open, centrifuge at maximum speed for 5 minutes. Centrifugation with an open lid prevents ethanol carry over during RNA elution and dries the spin column membrane. Drying is important because residual ethanol can interfere with downstream applications.
21. Discard the flow-through and collection tube.
22. Place the RNeasy MinElute spin column in a new 1.5mL collection tube.
23. Add the appropriate amount of RNase-free water or buffer for your library prep kit to the centre of the spin-column membrane (22µL RNase-free water).

Close the lid, and then centrifuge at maximum speed for 1 minute. Expect to recover ~18-20µL RNA.

### **2.11 RNA quantification**

The quality and quantity of total RNA was assessed using three methods: Nanodrop spectrometer (Thermo), Qubit assay kit (Thermo), and the Bioanalyzer system (Agilent). Nanodrop spectrometer (Thermo) provides quick quantification results and wavelength ratios that are useful in identifying contaminants. The Qubit RNA assay kit (Thermo) is a more sensitive method that uses fluorescence to selectively quantify low-abundance RNA samples. Qubit quantification was done with the broad range RNA (BR) reagents kit. The Bioanalyzer system (Agilent) is a chip-based capillary electrophoresis method that was used to assess the size and quality of nucleic acids over a range of sizes and concentrations. The bioanalyzer provides an RNA integrity number (RIN), electrophoresis gel images, and an electropherogram of nucleic acid peaks. RIN is a measure of RNA quality that is estimated by the Bioanalyzer instrument (Schroeder *et al.* 2006). The RNA 6000 Nano (Agilent) reagents were used, and RNA analysis was performed using Prokaryote Total RNA Series II module which quantifies total RNA levels and confirms presence of prokaryotic RNA in samples by showing bacterial 16S and 23S ribosomal fragments.

### 2.12 Quality control, mapping and counting RNA-Seq reads

The quality of raw FASTQ reads were checked by FastQC v0.11.5, Babraham Bioinformatics, UK (S. 2010) and visualised using multiQC v1.9 (Ewels *et al.* 2016). Trimmomatic v0.39 (Bolger, Lohse, and Usadel 2014) was used to removed residual TruSeq adapter sequences and low-quality bases. Reads with Phred quality scores <20 within a 5bp sliding window were dropped. Reads with total length <40bp were also dropped. FastQC and multiqc were run again to examine results of trimming and to confirm adapter sequence removal. Trimmed reads were then mapped to reference genomes using bowtie2 with default settings (Langmead and Salzberg 2012). Conversion into BAM files was performed using SAMtools (Li *et al.* 2009). Mapped reads were visualized in the Integrated Genome Browser (IGB) (Nicol *et al.* 2009). FeatureCounts v2.0.0 was used to summarize read counts for each annotated feature in the respective reference genomes in multimapping mode (-M) (Liao, Smyth, and Shi 2014).

### 2.13 Differential gene expression analyses

The count matrix was imported into R-studio (R v3.4.2) on a MacBook Pro (16-inch, 2019, v11.5.2). Transcripts with read counts <10 across all samples were filtered and removed prior to analysis. The Bioconductor package DESeq2 (v1.32.0) pipeline was used to normalise samples for sequencing depths and transcripts read length and to identify differentially expressed genes (DEGs) between conditions (Love, Huber, and Anders 2014). DESeq2 (v1.32.0) robustly tests for differential expressions using a negative binomial generalised linear model (Schurch *et al.* 2016). DESeq2 normalized libraries were regularized log transformed for visualizing heatmaps and clustering. Differential gene expression was performed on raw counts using transcripts per million (TPM). Features (genes) with a fold change >1.5 and false discovery rate (FDR) of <0.05 were categorized as differentially expressed, with genes showing increased expression relative to controls termed “upregulated” and genes with reduced expression relative to the controls termed “downregulated”. The terms “upregulated” and “downregulated” were used throughout this thesis to indicate direction of gene expression.

### **2.14 Biological pathways analysis**

To reveal functional properties of the differentially expressed genes in the conditions studied, pathway enrichment analysis was performed using the KEGG database annotation for genomes of serotypes which were the same as study strains (Goeman and Buhlmann 2007; Kanehisa 2000). Pathway enrichment analysis was performed separately for upregulated and downregulated genes. Pathway over-representation analysis (ORA) of DEGs was performed in R (v3.4.2) with clusterProfiler v4.0.5 (Yu *et al.* 2012). ORA results with and adjusted p-value (FDR) <0.05 were categorised as significantly enriched. The reported p-values were adjusted for multiple testing using the Benjamini-Hochberg methods.

### **2.15 Statistical analysis**

All statistical analysis were carried out in R-studio (R v3.4.2) using Bioconductor or CRAN packages. The specific statistical methods used are described in each chapter. All graphs presented in in this thesis were generated with ggplot2 R package (v3.3.6).

## Chapter 3 : Growth curves and optimisation experiments

### 3.1 Introduction

RNA sequencing (RNA-Seq) experiments are aimed at capturing cellular expression levels, under specific environments, and at a specific timepoint (Croucher and Thomson 2010). Therefore, careful experimental design is imperative for reproducibility of transcriptomic experiments but also to ensure that underlying biology of interests are being captured. Several environmental factors can influence bacterial growth and metabolism which need to be tightly controlled in transcriptomics studies. Environmental conditions that have been shown to influence bacterial behaviour include nutrients levels (e.g., carbohydrates source), bacterial growth phase, growth temperature, pH, etc (Kroger *et al.* 2013; van Opijnen and Camilli 2012; Caglar *et al.* 2017). Under experimental settings, the primary nutrient source for bacterial growth is typically the culture medium (Sanchez-Rosario and Johnson 2021). In *S. pneumoniae*, the choice of growth media has considerable influence on pneumococcal growth (Sanchez-Rosario and Johnson 2021). The two main classes of media used for *S. pneumoniae* growth are enriched media (e.g., THY, BHI) and chemically defined media (CDM) (Sanchez-Rosario and Johnson 2021). The two main enriched media types used for pneumococcal growth are THY and BHI which are designed to support optimal growth of broad range of bacterial species (Todd and Hewitt 1932; Sanchez-Rosario and Johnson 2021). Enriched media are usually commercially available in powder formulations and are rich in carbohydrates, amino acids, transition metal ions and vitamins to enable maximal bacterial growth (Todd and Hewitt 1932; Sanchez-Rosario and Johnson 2021). CDM on the other hand are designed to enable altering constituents of the media making it possible to assess the specific influence of different components of the medium (Sanchez-Rosario and Johnson 2021). Bacterial growth in CDM is usually more restricted compared to enriched media. Bacterial growth phase is also an important influence on the microbial transcriptome which should be carefully considered and controlled for (Caglar *et al.* 2017). As bacterial pathogens grow through the different growth phases they quickly adapt and refine cellular transcriptome (Caglar *et al.* 2017; Kroger *et al.* 2013). If not carefully controlled at the experimental stage, bacterial transcriptomic results may reflect normal cellular changes during bacterial growth rather than environmental exposures being evaluated (Caglar *et al.* 2017). A strategy to mitigate the influence of normal growth on transcriptome is to use short exposure times to evaluate experimental conditions of interest (Kroger *et al.* 2013).

The aim in this chapter was to identify optimal growth and experimental approaches for generating high-quality transcriptomics data of *S. pneumoniae*. In addition, the growth characteristics of *S. pneumoniae* strains were evaluated under various environmental conditions.

### 3.2 Results

#### 3.2.1 Optimisation experiments

##### 3.2.1.1 *S. pneumoniae* growth in 7H9+ media

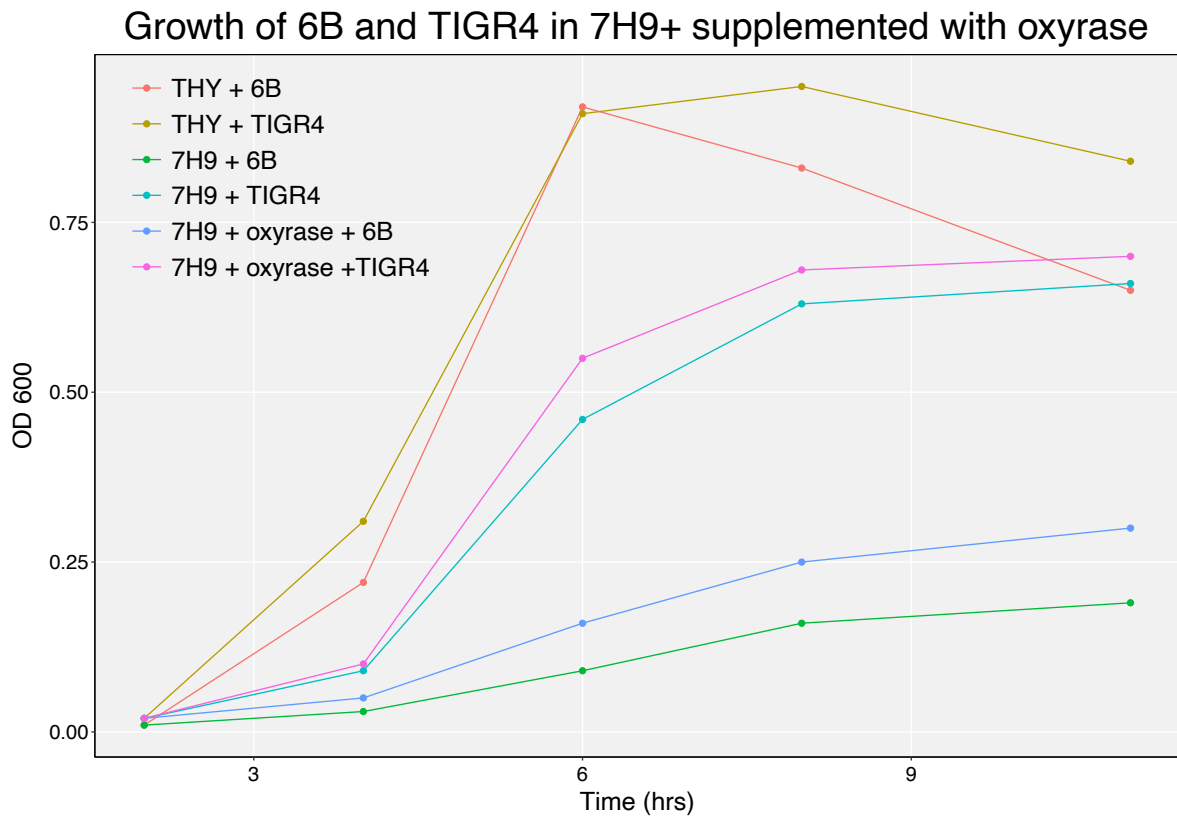
The 7H9+ media was developed by Dr Lavanya Mane to investigate the metabolic profile of *S. pneumoniae* serotype 4 (TIGR4) strain and *Mycobacterium tuberculosis* (MTB) including in some of the environmental conditions being evaluated in this thesis (unpublished). The use of 7H9+ media for the proposed work in this thesis would have presented an opportunity to exploit transcriptomics (from this thesis) and metabolomics (from Dr Mane's work) datasets to understand *S. pneumoniae* response to specific environmental conditions at multi-omics level. Therefore, 7H9+ was the media of choice from the onset to generate transcriptomic data in this thesis. As a result, the initial "higher invasive" clinical pneumococcal strain chosen for the work in this thesis was the TIGR4 (serotype 4) strain, similar to Dr Mane, but this was later changed to a serotype 1 strain (BVJ1JL). The serotype 6B (BHN418) strain was chosen in addition because this will enable *in vivo* validation of transcriptomic results from the EPHC model. Multiple growth evaluation experiments of the serotype 6B strain however exhibited very slow growth, reaching OD<sub>600</sub> of less than 0.2 after 12 hours in 7H9+ media. In the same growth conditions, the serotype 4 (TIGR4) strain grew to a maximum OD of at least 0.5 with exponential phase occurring between 3 – 7 hours after incubation (**Figure 3.1**).

With the aim of improving the growth of serotype 6B in 7H9+ media, several modifications were tested. First, growth of serotype 6B in 7H9+ medium not supplemented with tyloxapol was tested. Tyloxapol is a detergent that was specifically added in the original media formulation to support growth of MTB. Visual inspection of cultures of serotype 6B grown in 7H9+ suggested inhibition of growth by tyloxapol with more than 10 times fewer

viable colonies observed on CBA in tyloxapol supplemented cultures compared to non-supplemented media formulation. However, this did not result in faster growth, as measure by OD<sub>600</sub> (**Table 3.1**). Using of a higher starting bacterial inoculum did not also enhance bacterial growth (**Table 3.1**). Next, it was evaluated whether a different carbohydrate source could enhance growth. The base carbohydrate in 7H9+ was glucose and this was substituted with galactose. Galactose as carbohydrate source reduced the growth rate of *S. pneumoniae* serotype 4 (TIGR4) and only slightly increased the growth rate of serotype 6B (BHN418) strain (**Figure 3.2**). Finally, addition of components of other CDMs to boost growth of serotype 6B was tested. Oxyrase, a component of the Chang CDM formulation, enhanced the growth rate and maximum growth densities of both the serotype 6B and serotype 4 strains (**Figure 3.1**). However, this growth enhancement was marginal for both strains. Furthermore, oxyrase is an oxygen-scavenging compound that removes oxygen from liquid media and would likely affect the bacterial transcriptome independent of environmental exposures being tested. Supplementation of 7H9+ medium with the amino acid formulations of the NMC media was tested, but this also did not boost growth of serotype 6B in 7H9+ media.

The Chang CDM supported growth of both serotype 6B (BHN418) and serotype 4 (TIGR4) strains but was not evaluated further because of the need to supplement with oxyrase, making the medium unsuitable for transcriptomics experiments.

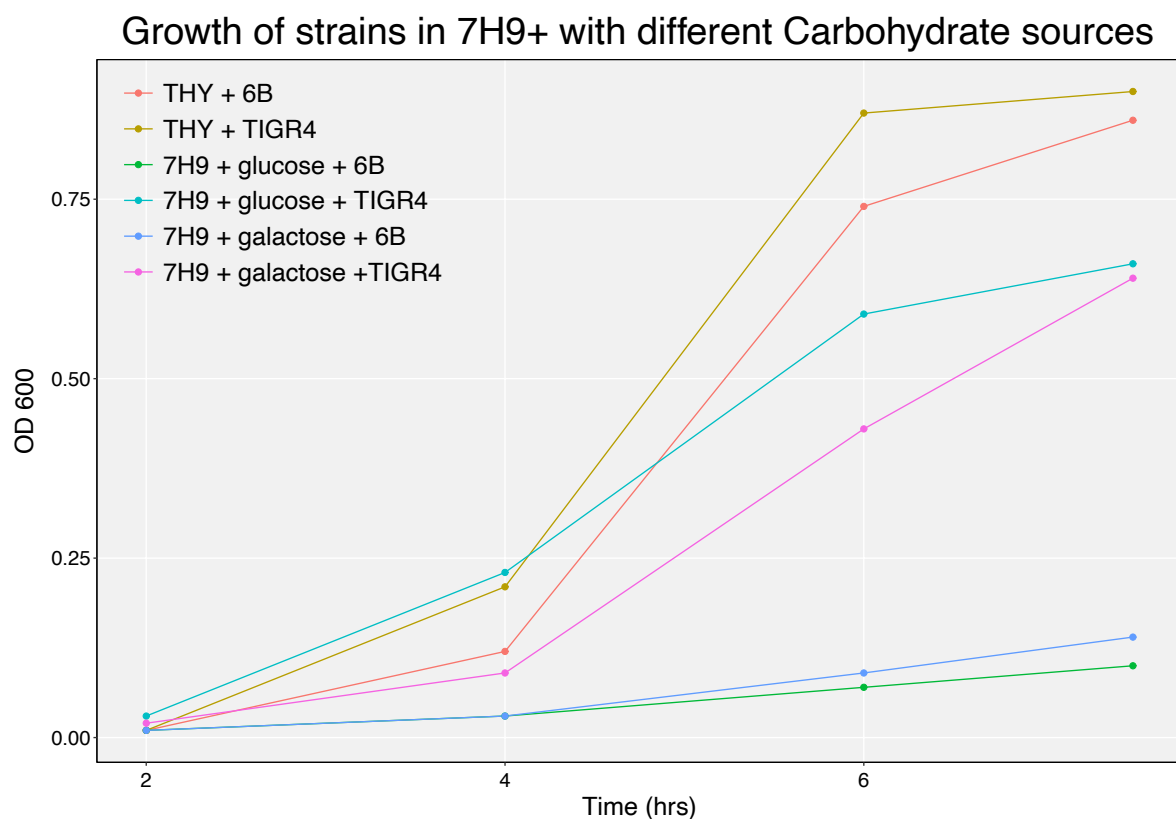




**Figure 3.1:** Evaluation of the effect of oxyrase – an oxygen scavenging compound on growth of *S. pneumoniae* serotype 6B (BHN418) and serotype 4 (TIGR4) strains in 7H9+ media. Serotype 6B grew poorly in 7H9+ media and was slightly boosted by addition of oxyrase. Serotype 4 (TIGR4) grew quicker and to a higher maximum density than serotype 6B (BHN418) in 7H9+ and this was further boosted by the addition of oxyrase. THY was used as control showing both serotype 6B (BHN418) and serotype 4(TIGR4) strains grew equally well. Experiments were conducted in 50mL falcon tubes incubated at 37°C with 5% CO<sub>2</sub> (n = 1).

**Table 3.1:** The effect of phosphate buffered saline (PBS) washing, addition of tyloxapol, and different concentrations of bacterial starting inoculum on *S. pneumoniae* (serotype 6B) growth in 7H9+. Growth in enriched media THY used as control. Experiments were performed in 50mL falcon tubes and incubated at 37°C with 5% CO<sub>2</sub> (n = 1.)

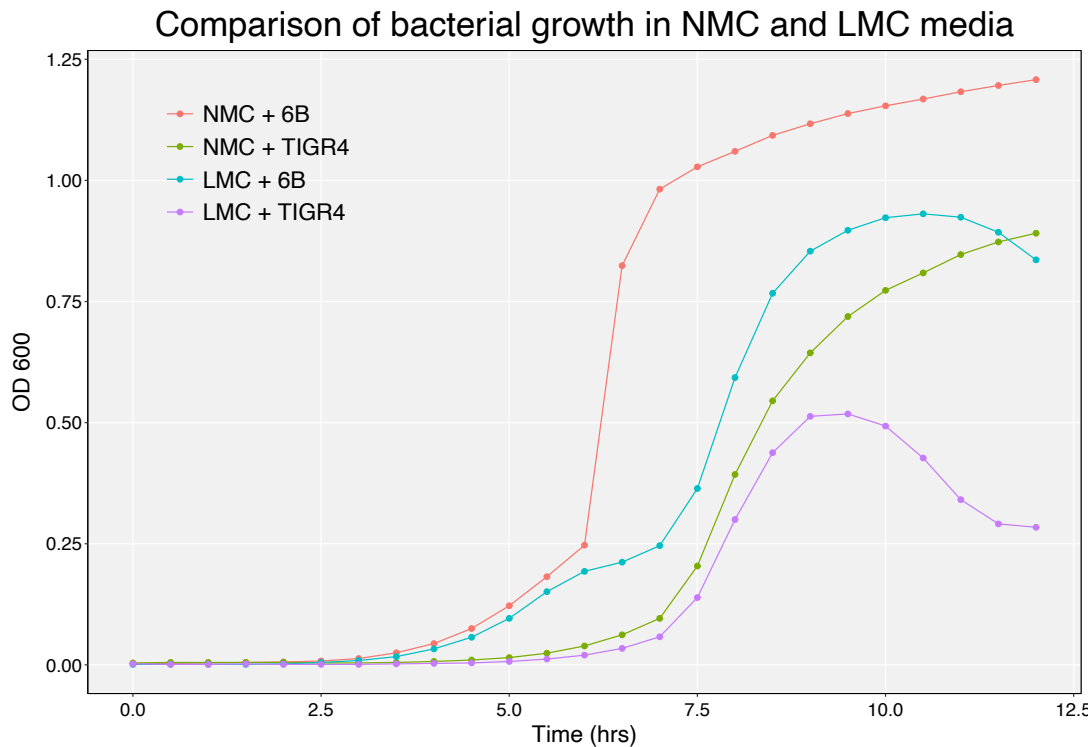
Condition	OD <sub>600</sub> (12hrs)	OD <sub>600</sub> (36hrs)	OD <sub>600</sub> (144hrs)
7H9+ with tyloxapol + 250µl bacteria (washed)	0.00	0.00	0.14
7H9+ without tyloxapol + 250µl bacteria (washed)	0.00	0.00	0.45
7H9+ without tyloxapol + 500µl bacteria (washed)	0.00	0.00	0.44
7H9+ without tyloxapol + 750µl bacteria (washed)	0.00	0.00	0.44
7H9+ without tyloxapol + No bacteria added	0.00	0.00	0.00
THY + 250uL bacteria (washed)	0.75	0.47	0.36
7H9+ without tyloxapol + 500µl bacteria (unwashed)	0.07	0.36	0.43
7H9+ without tyloxapol + 250µl bacteria (unwashed)	0.01	0.46	0.63



**Figure 3.2:** Substitution of glucose with galactose as carbohydrate source in 7H9+ did not significantly enhance growth of serotype 6B (BHN418). Serotype 4 (TIGR4) grew slower with galactose as carbohydrate source. Enriched media THY was used as control. Experiments conducted in 50mL falcon tubes incubated at 37°C with 5% CO<sub>2</sub> (n = 2).

### 3.2.1.2 Improved growth of both serotype 6B (BHN418) and serotype 4 (TIGR4) in NMC

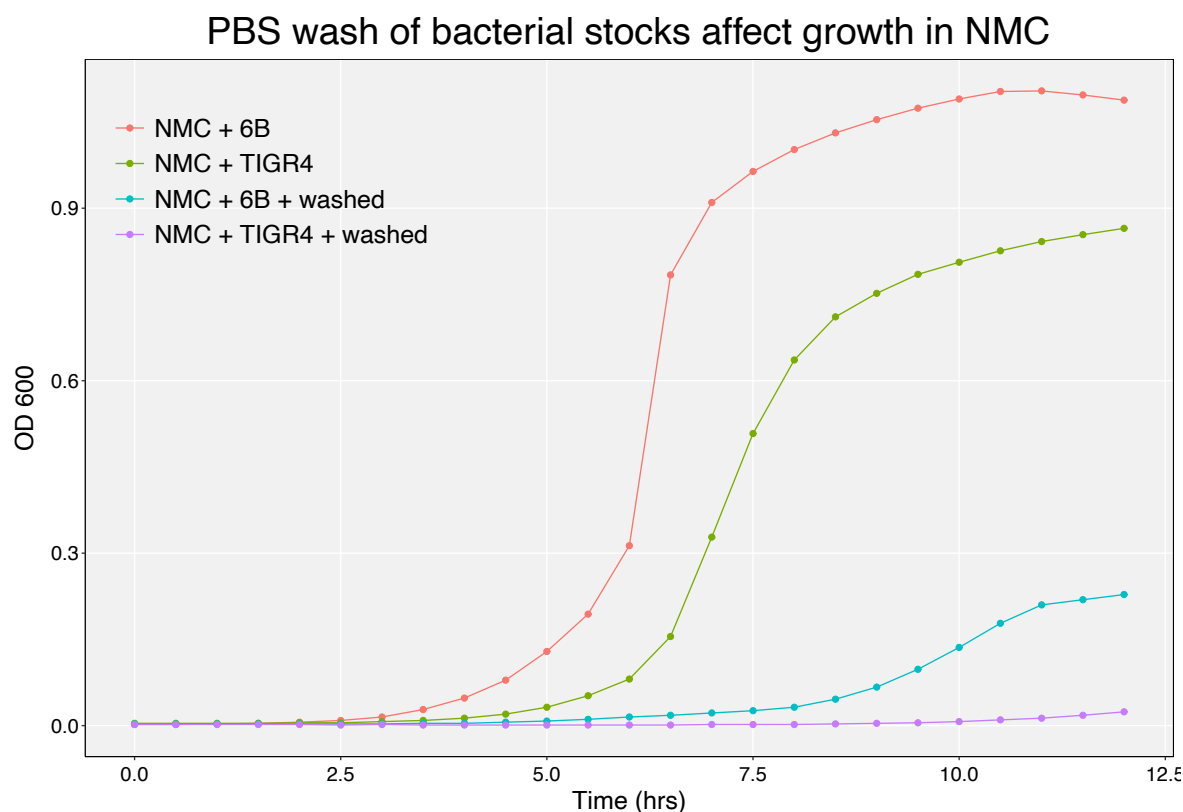
Having shown that 7H9+ and Chang CDM were not suitable growth media for transcriptomic experiments, the NMC and LMC media described by Aprianto *et al* (2018) were tested. Both NMC and LMC supported the growth of serotype 6B (BHN418) and serotype 4 (TIGR4) (**Figure 3.3**). The NMC was selected as the media of choice over LMC for subsequent transcriptomics experiments in this thesis because it was modelled by Aprianto *et al* (2018) to mimic the nasopharyngeal environment (colonisation niche).



**Figure 3.3:** Growth comparison of serotype 4 (TIGR4) and serotype 6B (BHN418) in lung mimicking condition (LMC) and nose mimicking condition (NMC) media. Experiments conducted with plate reader and incubated at 37°C with 5% CO<sub>2</sub> (n = 2). NMC media supported growth of both strains better as indicated by higher growth density compared to LMC.

### 3.2.1.3 Washing bacterial pellets with PBS inhibits growth

It is common to resuspend pellets of bacterial stocks from freezer storage in phosphate buffered saline (PBS) prior to inoculation to remove excess media. This washing step was found to considerably affect the growth of both serotype 6B (BHN418) and serotype 4 (TIGR4) strains in NMC compared to non-washed bacteria (**Figure 3.4**). PBS washing also negatively affected *S. pneumoniae* growth in 7H9+ media (**Table 3.1**). It was suspected that the washing process with PBS may have stressed the bacteria, leading to reduced viability when reinoculated in growth media. Therefore, subsequent growth curves were carried out directly from bacterial aliquots stored in the freezer.

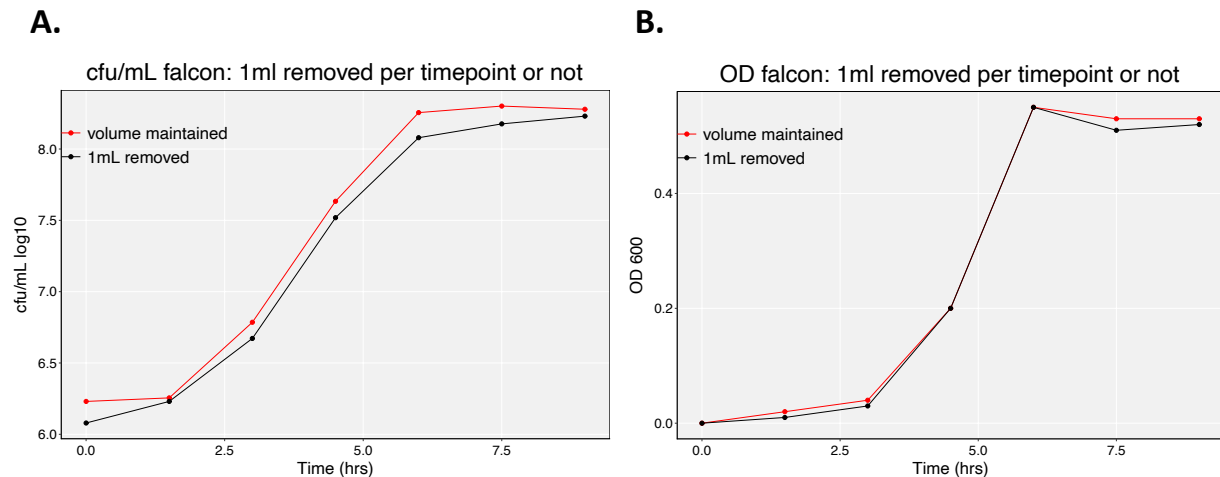


**Figure 3.4:** The effect of washing bacterial cells stored in glycerol with phosphate buffered saline (PBS) was evaluated. Washing bacterial stocks with PBS prior to inoculation in NMC negatively impacted bacterial growth as measured by OD<sub>600</sub>.

#### 3.2.1.4 Optical density of bacterial cultures correlated with amount of bacteria

Optical density (OD) measurement of cultures was used as proxy for the bacterial growth and the number of live bacteria in the culture. The selection of a specific growth point was important for transcriptomics experiments to ensure data reproducibility. In addition, although some of the optimisation growth curves were assessed with an automated plate reader platform, transcriptomic experiments would have to be conducted using manual methods in falcon tubes or flasks, which involved periodical checking of bacterial growth to assess whether cultures have reached a specific OD. Therefore, comparison experiments were conducted to investigate the correlation between culture OD with number of live bacterial as measured by CFU/mL and also to evaluate whether discarding culture volumes periodically used to measure OD affected this relationship. In one experimental set, the 1mL volume removed from cultures to measure OD and calculate bacterial CFU/mL was discarded. In the other experiment, 0.8mL volume was return to

original culture and re-incubated (0.2mL used for CFU/mL enumeration). The results of this experiment showed that there was a high correlation between bacterial culture OD and CFU/mL values. Periodic removal of 1mL culture did not influence this relationship (Figure 3.5).



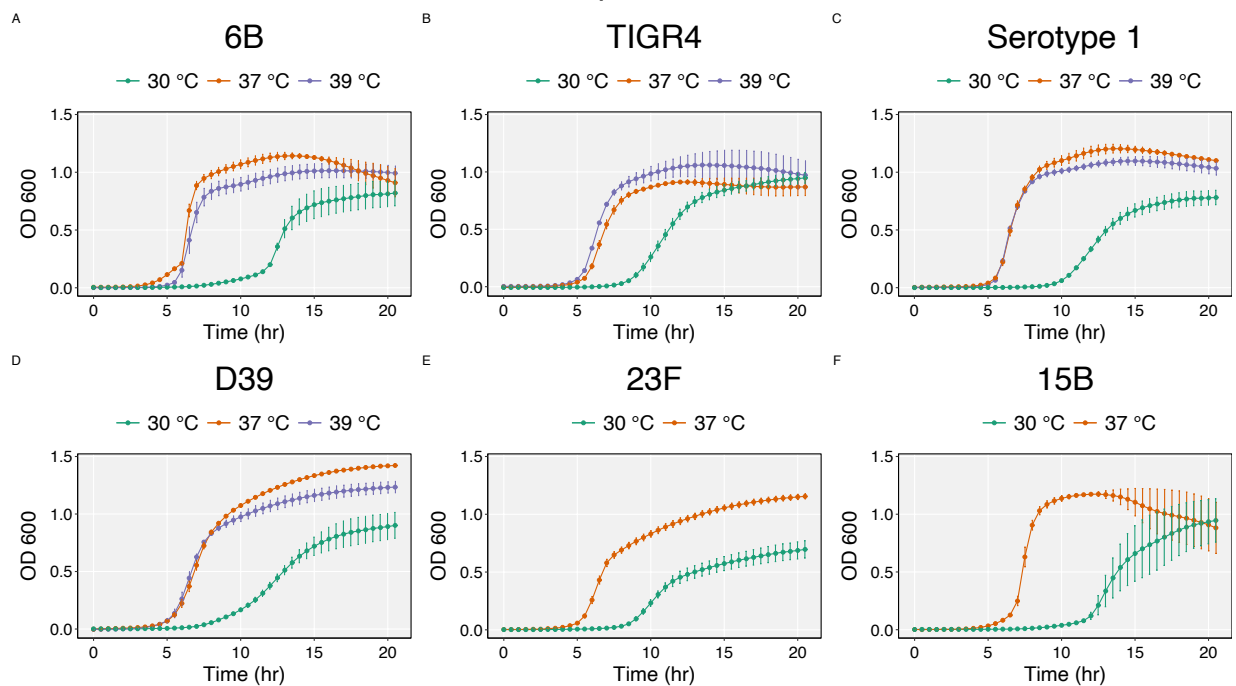
**Figure 3.5:** Optical density (OD) as measured by spectrometer correlates well with CFU/mL counts (Pearson  $r = 0.961$ ). Discarding 1mL of bacterial suspension used to measure OD does not affect subsequent OD values or CFU/mL concentration. Pneumococcal strain BHN418 (serotype 6B) was used in this experiment.

### 3.2.2 Comparison of growth of *S. pneumoniae* strains in different growth conditions

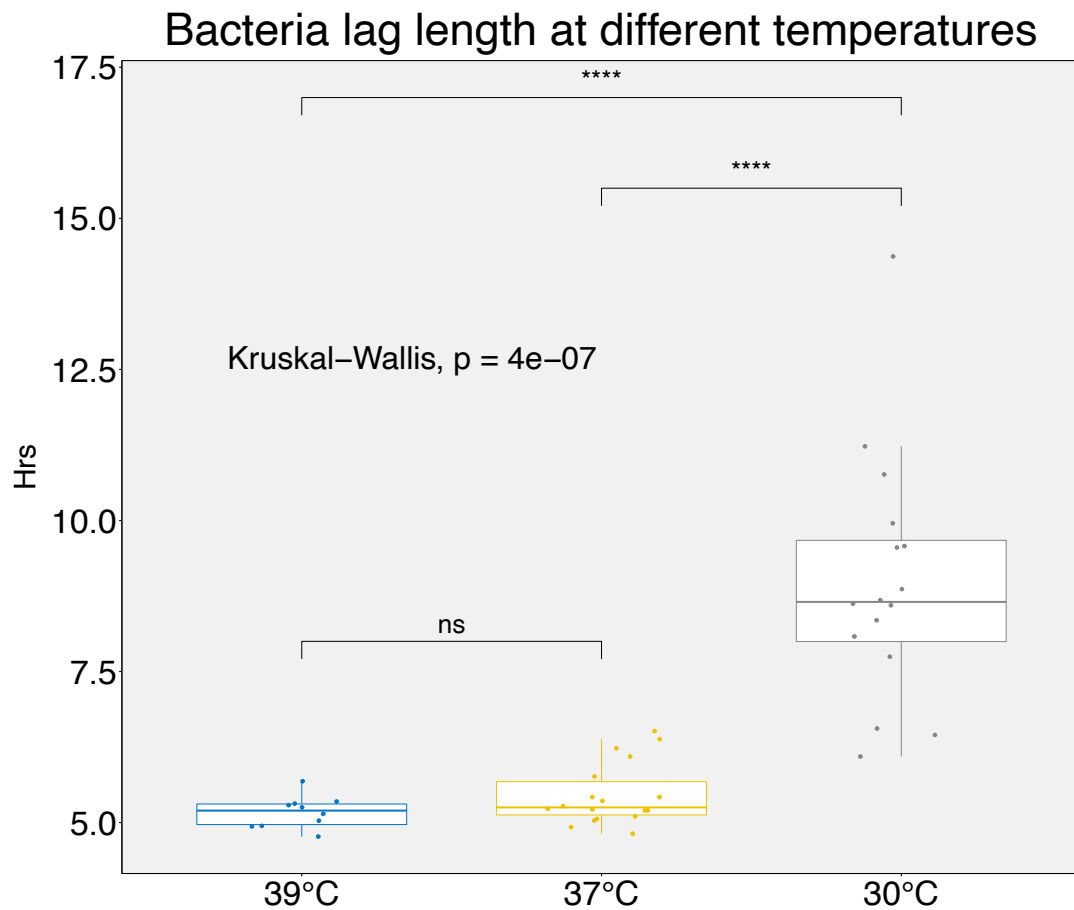
#### 3.2.2.1 Growth parameters of *S. pneumoniae* were influenced by temperature

*S. pneumoniae* strains of BHN418 (serotype 6B), TIGR4 (serotype 4), BVJ1JL (serotype 1), D39 (serotype 2), P1262 (serotypes 15B) and P1121 (serotype 23F) grew exponentially at 37°C and 39°C (Figure 3.6). However, growth at 30°C resulted in slower growth rate, longer lag phase, and longer bacterial doubling time for all strains compared to growth at 37°C and 39°C (Figure 3.6 and Table 3.2). Higher temperatures (39°C and 37°C) resulted in faster growth rate (Figure 3.6) and maximum growth density of *S. pneumoniae* compared to growth at 30°C. This growth pattern was consistent across various *S. pneumoniae* serotypes grown in NMC (Figure 3.6 and Figure 3.7).

## Growth kinetics of strains at different temperatures



**Figure 3.6:** Growth of different *S. pneumoniae* serotypes at temperatures of 30°C, 37°C, and 39°C. The means of independent experiments ( $n = 3$ ) is shown. The error bars represent standard error from mean (SEM) between experiments. For each experiment, 3 technical replicates were included.



**Figure 3.7: *S. pneumoniae* growth at 30°C is significantly different from growth at 37°C and 39°C.** Boxplot showing the calculated length of lag growth phase for all strains (serotype 6B, serotype 4, serotype 1, serotype 2, serotype 23F, and serotype 15B) at 30°C, 37°C, and 39°C. Pneumococcal growth at 30°C was significantly delayed for all strains compared to growth at 37°C and 39°C. Statistical analysis using Kruskal-Wallis was conducted in R for group comparisons.



**Table 3.2:** Growth characteristics of different *S. pneumoniae* serotypes at 30°C, 37°C, and 39°C in NMC media. Experiments were performed using a plate reader. Values were the mean of independent experiments (n=3) showing standard error from mean (SEM). Growth parameters were calculated using the curveball algorithm (Ram *et al.* 2019).

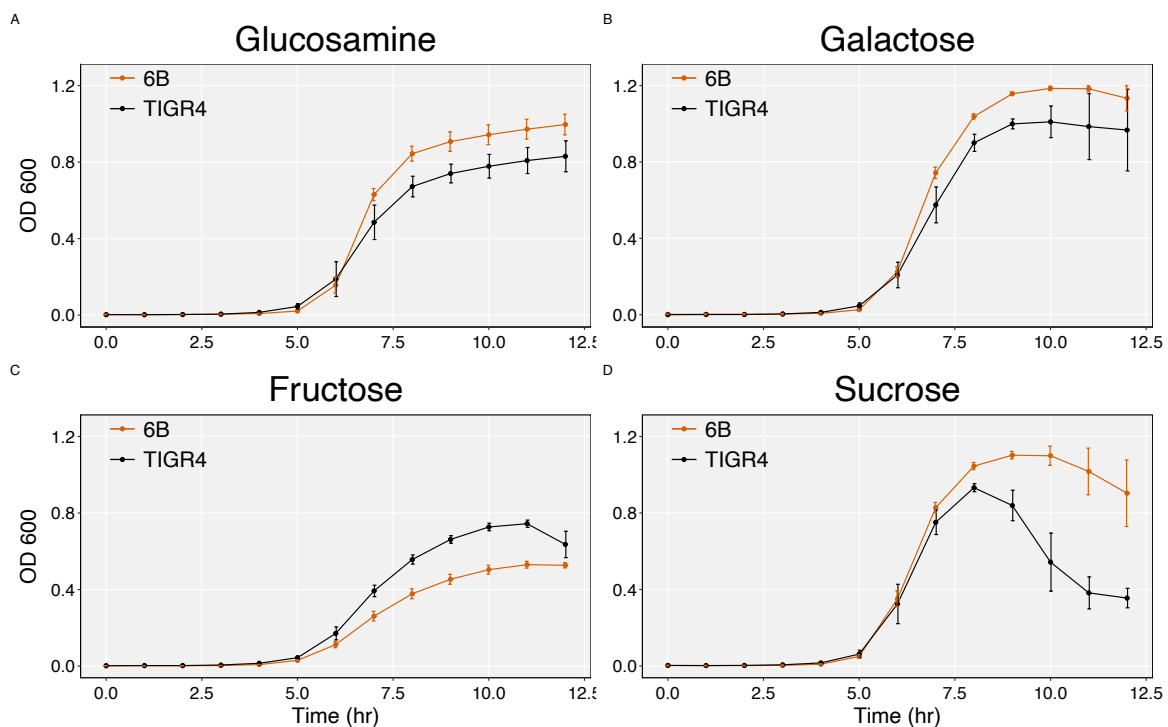
Serotype (strain)	Growth temperature	Lag length (hrs)	Growth rate (hr <sup>-1</sup> )
6B (BHN418)	30°C	8.45 (± 0.20)	0.91 (± 0.47)
1 (BVJ1JL)	30°C	9.87 (± 0.29)	0.67 (± 0.02)
4 (TIGR4)	30°C	8.54 (±0.38)	0.76 (±0.03)
15B (P1262)	30°C	11.50 (± 0.39)	1.11 (±0.01)
2 (D39)	30°C	8.94 (± 0.55)	0.63 (± 0.02)
23F (P1121)	30°C	7.77 (±0.30)	0.78 (± 0.17)
6B (BHN418)	37°C	5.19 (± 0.09)	1.76 (± 0.05)
1 (BVJ1JL)	37°C	5.66 (± 0.63)	3.24 (± 1.20)
4 (TIGR4)	37°C	5.28 (± 0.12)	1.87 (± 0.43)
15B (P1262)	37°C	6.28 (± 0.21)	3.11 (± 0.67)
2 (D39)	37°C	5.41 (± 0.35)	1.54 (± 0.16)
23F (P1121)	37°C	4.93 (± 0.15)	2.05 (± 0.85)
6B (BHN418)	39°C	5.50 (± 0.26)	4.44 (± 0.15)
1 (BVJ1JL)	39°C	5.27 (± 0.03)	3.99 (± 1.14)
2 (D39)	39°C	5.06 (± 0.41)	1.49 (± 0.05)
4 (TIGR4)	39°C	4.99 (± 0.06)	2.46 (± 1.28)

### 3.2.2.2 Growth in different carbohydrate sources

N-acetylglucosamine, galactose, sucrose, and fructose as carbohydrate sources supported the growth of *S. pneumoniae* serotype 6B (BHN418) and serotype 4 (TIGR4) in NMC. The growth kinetics of the strains with each of the four carbohydrate sources substituted in the media is shown in **Figure 3.8**. All four carbohydrate sources supported exponential growth of strains to a maximum OD<sub>600</sub> of at least 0.8 except for fructose (**Figure 3.8**). In fructose, serotype 4 (TIGR4) was the faster growing strain (twice as fast as serotype 6B) with a higher maximum OD<sub>600</sub> in contrast to other carbohydrate sources tested. Serotype

6B (BHN418) had the fastest growth rate with N-acetylglucosamine, the base carbohydrate of the NMC media and the most abundant carbohydrate source in the nasopharyngeal niche (**Figure 3.8 and Table 3.3**). Sucrose induced the fastest growth rate for serotype 4 (TIGR4) (**Table 3.3**) but this was accompanied with the quickest autolysis rate compared to the other carbohydrate sources evaluated (**Figure 3.8**). Since sucrose is a disaccharide containing both glucose and fructose, the increase growth for both strains in sucrose compared to fructose could be attributed to the glucose component of sucrose. The length of lag phase between the strains was similar with the different carbohydrate sources tested (**Table 3.3**).

### Growth of 6B and TIGR4 in Different Carbohydrates Sources



**Figure 3.8:** Growth kinetics of *S. pneumoniae* serotype 6B (BHN418) and serotype 4 (TIGR4) in four different carbohydrates sources. The graphs show the data from independent experiments (n=3). For each experiment at least 3 technical replicates were included. The error bars represent SEM between the independent experiments.

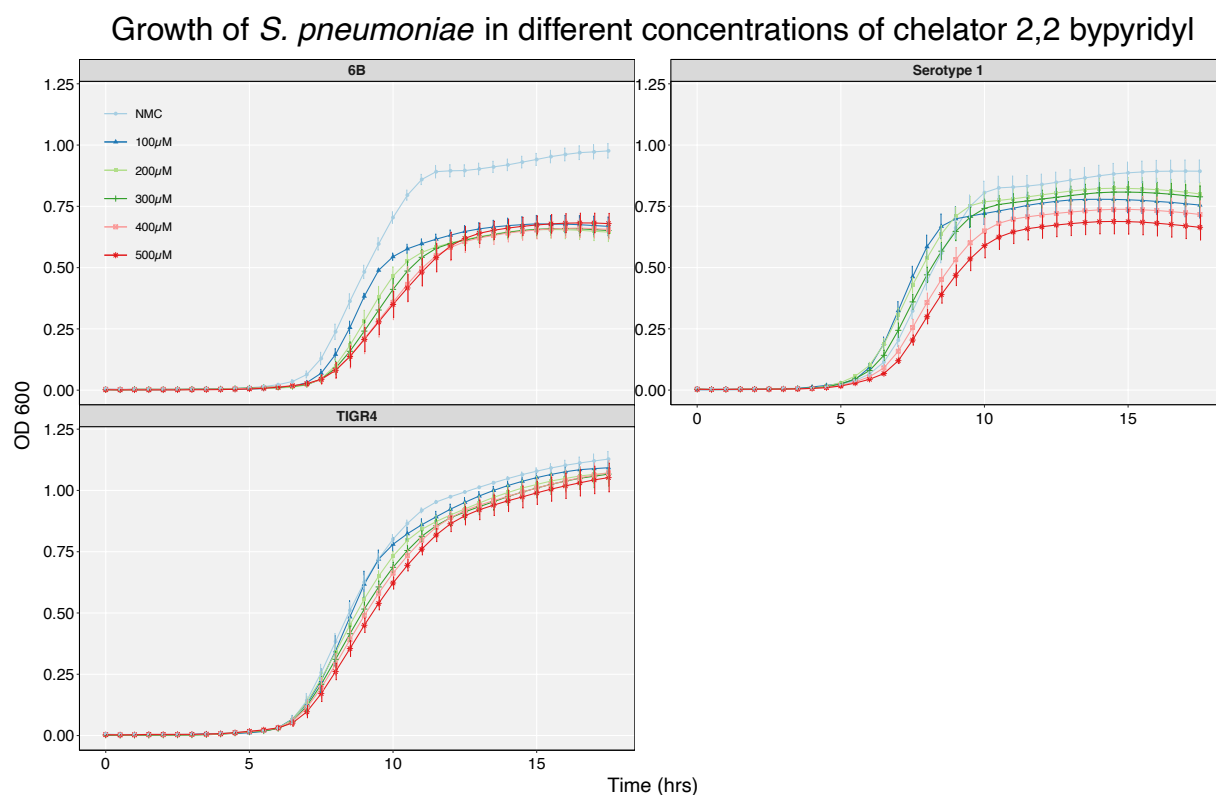
**Table 3.3:** Growth parameters of *S. pneumoniae* serotype 6B (BHN418) and serotype 4 (TIGR4) during growth in N-acetylglucosamine, galactose, fructose, and sucrose. Independent experiments (n=3) were generated in plate reader incubated at 37°C with 5% CO<sub>2</sub>. Growth parameters were calculated using the curveball algorithm (Ram *et al.* 2019).

Serotype	Carbohydrate	Lag length (hrs)	Growth rate (hr <sup>-1</sup> )
6B	N-acetylglucosamine	5.30 (± 0.02)	3.57 (± 0.40)
4	N-acetylglucosamine	5.36 (± 0.04)	2.13 (± 0.18)
6B	Galactose	5.42 (± 0.05)	2.25 (± 0.01)
4	Galactose	5.36 (± 0.27)	1.62 (± 0.27)
6B	Fructose	5.11 (± 0.15)	1.17 (± 0.03)
4	Fructose	5.04 (± 0.18)	1.89 (± 0.03)
6B	Sucrose	5.07 (± 0.07)	2.43 (± 0.15)
4	Sucrose	5.07 (± 0.23)	2.73 (± 0.24)

### 3.2.2.3 Effect of transition metals concentrations on *S. pneumoniae* growth

Transition metals chelator 2,2 bipyridyl (BIP) was used at concentrations ranging from 1µM – 500µM to assess the effect of transition metals stress on growth of *S. pneumoniae* serotype 1 (BVJ1JL), serotype 6B (BHN418), and serotype 4 (TIGR4) strains. Increasing doses of the chelator led to longer lag phase and increased doubling time for all strains at BIP concentrations 400µM-500µM (**Figure 3.9**). Unique to serotype 1, growth at the lower concentrations of 100µM, 200µM, and 300µM of the chelator resulted in a faster growth rate and shorter lag phase, although maximal growth density was lower compared to the control (**Figure 3.9** and **Table 3.4**). The concentration of transition metals in NMC has not been measured in this thesis and was not reported by Aprianto *et al* (Aprianto *et al.* 2018). To validate the chelating efficiency of BIP in NMC and its effect on *S. pneumoniae* growth, we used enriched media THY and serotype 6B at even lower BIP concentrations ranging from 1µM – 300µM. THY contains 740µg l<sup>-1</sup> iron, up to four-fold higher than in a CDM (Hoyer *et al.* 2018). We observed a similar effect of BIP on growth of strains in THY with increasing concentrations of BIP 1µM – 300µM resulting in longer lag phase and slower

growth rate (**Figure 3.10**). Growth parameters for serotypes 1 (BVJ1JL), serotype 6B (BHN418), and serotype 4 (TIGR4) in NMC are shown in **Table 3.4**.

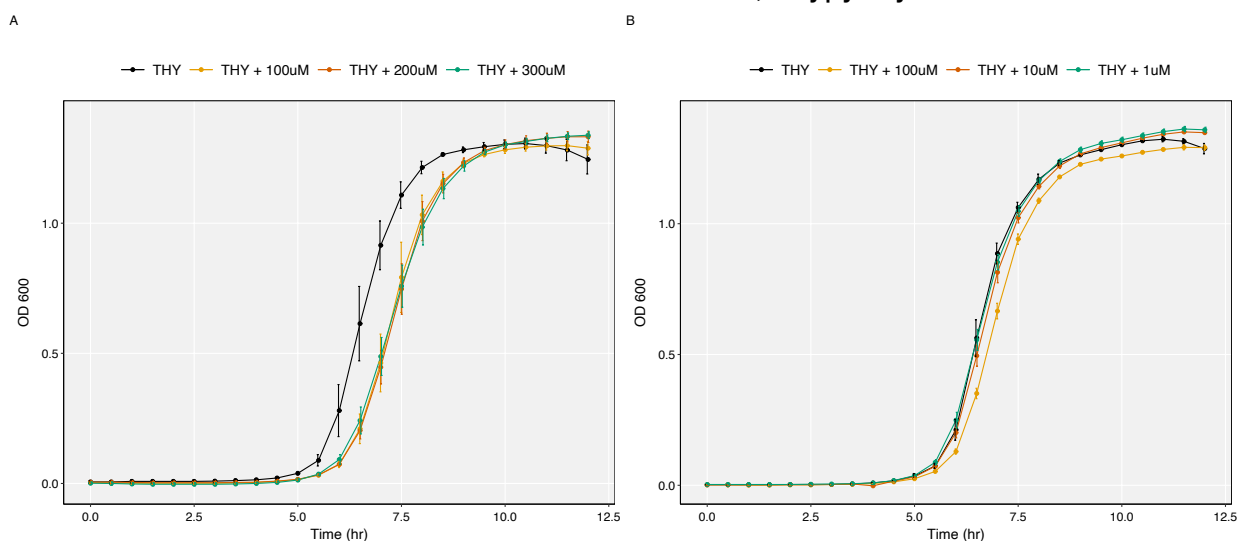


**Figure 3.9:** Growth of *S. pneumoniae* serotype 6B (BHN418), serotype 4 (TIGR4), and serotype 1 (BVJ1JL) in varying concentration of iron chelator 2,2 bipyridyl (BIP) cultured in NMC media. Data presented are average values with error bars representing SEM obtained in triplicate experiments (n=3). Increased concentrated of BIP generally increased length of lag phase and reduced maximum growth density.

**Table 3.4:** Growth parameters of *S. pneumoniae* serotype 6B (BHN418), serotype 4 (TIGR4), and serotype 1 (BVJ1JL) strains in NMC supplemented with different concentration of iron chelator 2,2, bipyridyl. OD<sub>600</sub> was measure from independent experiments (n=3) with plate reader incubated at 37°C with 5% CO<sub>2</sub>. SEM between experiments is in parentheses.

Serotype	Concentration of 2,2 bipyridyl (μM)	Lag length (hrs)	Growth rate (hr <sup>-1</sup> )
6B	0	7.02 (± 0.16)	1.42 (± 0.18)
4	0	6.30 (± 0.14)	0.96 (± 0.03)
1	0	6.05 (± 0.07)	1.11 (± 0.02)
6B	100	7.16 (± 0.13)	1.49 (± 0.25)
4	100	6.34 (± 0.22)	0.95 (± 0.03)
1	100	5.63 (± 0.07)	1.4 (± 0.11)
6B	200	7.47 (± 0.25)	1.24 (± 0.27)
4	200	6.3 (± 0.15)	0.92 (± 0.02)
1	200	5.65 (± 0.13)	1.24 (± 0.04)
6B	300	7.6 (± 0.28)	1.43 (± 0.15)
4	300	6.38 (± 0.19)	0.89 (± 0.02)
1	300	5.83 (± 0.12)	1.33 (± 0.13)
6B	400	7.84 (± 0.40)	1.07 (± 0.08)
4	400	6.50 (± 0.26)	1.01 (± 0.13)
1	400	6.16 (± 0.13)	1.1 (± 0.05)
6B	500	7.83 (± 0.35)	0.96 (± 0.04)
4	500	6.59 (± 0.22)	0.86 (± 0.01)
1	500	6.32 (± 0.05)	1.04 (± 0.02)

## Growth of 6B in THY with various concentration of 2,2 bipyridyl



**Figure 3.10:** Growth of *S. pneumoniae* serotype 6B (BHN418) in THY supplemented with different concentrations of 2,2, bipyridyl (BIP). Mean OD<sub>600</sub> from independent experiments (n=3). Increasing concentration of BIP resulted in longer lag phase in a dose dependent manner in concentrations ranging from 1μM - 300μM in enriched media.

### 3.2.3 Assessment of RNA isolation protocol for high-quality *S. pneumoniae* RNA

The RNA isolation protocol was piloted by growing BHN418 (serotype 6B) in 25mL of NMC and THY incubated at 37°C with 5% CO<sub>2</sub> to mid-log phase in 200mL conical flasks. The RNA in these samples were stabilized by addition of RNALater (Ambion) according to manufacturer's protocol. As part of this assessment, the efficiency of DNA digestion and purification of total RNA yield in our RNA isolation protocol (Chapter 2) was evaluated. An additional RNA cleaning step with RNeasy mini-elute clean-up kit (Qiagen) together with DNA digestion of RNA samples yielded up to 13-fold more RNA (**Table 3.5**), although this resulted in some RNA degradation as indicated by the lower RNA integrity number (RIN) score in both THY and NMC (Schroeder *et al.* 2006). RNA purity was high as indicated by 260/280 and 260/230 nanodrop ratios of ~2 (**Table 3.5**). Due to the higher quality of RNA product that resulted from the DNA digestion and cleaned-up, these procedures were incorporated in RNA isolation protocols.

**Table 3.5:** Quantification results evaluating DNA digestion and extra RNA clean-up on quality of RNA isolation protocol.

Media	Description	Qubit		Nanodrop		Bioanalyzer		
		BR-RNA		260/280	260/230	ng/ $\mu$ L	RIN	23s/ 16s
		ng/ $\mu$ L	ng/ $\mu$ L					
THY	C	504	390	2.08	2.01	236	2.6	0.0
THY	D+ R+	2360	2366.3	2.15	2.34	1893	6.2	0.8
THY	D+ R-	296	405	2.08	1.99	410	7	0.9
THY	D- R-	141.8	390	2.15	0.88	399	6.5	1.0
NMC	C	Too low	2.5	3.03	0.03	0	0.0	NA
NMC	D+ R+	3400	3986.4	2.2	2.42	3221	5.8	0.7
NMC	D+ R-	195	286	2.09	2.02	299	6.5	0.7
NMC	D- R-	244	348.5	2.15	2.07	256	7	1.0

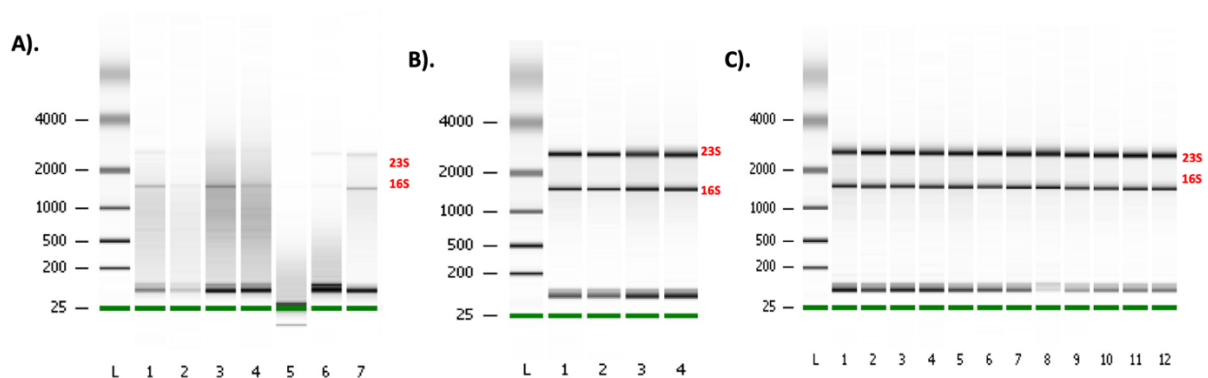
RIN = RNA integrity number

Description: C = control; D = DNA digestion; R = RNeasy Cleanup; + = done; - = not done

### 3.2.3.1 Phenol-ethanol stabilization degrades *S. pneumoniae* RNA

To ensure that transcriptomic experiments capture snapshots of the specific timepoints being investigated, stabilization of RNA to prevent continuing cellular transcription and protein processing was imperative. The phenol-ethanol “stop solution” used by Kroger *et al* (Kroger *et al.* 2013; Kroger *et al.* 2018) was the initial choice because of simplicity of preparation, cost considerations, experience of the author group regarding bacterial transcriptomics, and the high quality of data reported in previous studies using this method. However, multiple experiments (n=4) using the phenol-ethanol stabilization method resulted in degraded RNA with RIN scores of less than 4.1 in all experiments (**Figure 3.11**). The original studies that used the phenol-ethanol “stop-solution” were conducted for transcriptomic experiments on Gram-negative pathogens *Salmonella typhimurium* and *Acinetobacter baumannii* which may explain the difference in results. RNALater (Ambion) and RNA Protect Bacteria (Qiagen) resulted in intact RNA with both commercial stabilization reagents resulting in RIN scores ranging from 7.0 – 9.6. RNALater

was subsequently not considered as a suitable stabilization reagent because of the manufacturer's recommendation to obtain cellular pellets by centrifugation prior to the addition of stabilization reagent. Following this recommendation would imply that bacterial cellular transcription and RNA turnover will continue to occur after intended timepoints during centrifugation and thus may generate misleading transcriptomic data. Based on these results, RNA Protect Bacteria (Qiagen) was selected to conduct future transcriptomic experiments because it stabilizes bacterial RNA in growth media and would therefore arrest RNA transcription at the specifically selected timepoints.



**Figure 3.11:** Effects of RNA stabilization reagents A). Phenol-ethanol, B). RNALater (Ambion), and C). RNA Protect Bacteria (Qiagen) on RNA quality. Phenol-ethanol as a stabilization solution degraded *S. pneumoniae* RNA with RIN scores of <4.1. RNALater and RNA Protect Bacteria both achieved intact RNA with RINs of >8.0. L = ladder, 1-12 = Samples



### 3.3 Discussion

The optimisation experiments undertaken in this chapter were important for establishing robust *S. pneumoniae* comparative transcriptomics experiments. Through the work in this chapter, informed decisions towards experimental design and the generation of a high-quality RNA-seq dataset were made. It was established that among the different chemically defined media tested, NMC best supported the growth of *S. pneumoniae* in the environmental conditions of interest and would be the most suitable for transcriptomics experiments (Aprianto *et al.* 2018). An efficient protocol for the isolation of high-quality *S. pneumoniae* RNA was developed and refined. The RNA Protect Bacteria reagent from Qiagen was established as an optimal RNA stabilisation reagent for transcriptomics experiments of *S. pneumoniae*. Finally, growth curve analysis highlighted differential growth phenotypes of *S. pneumoniae* strains to changes in carbohydrate source, temperature, and concentration of transition metal ion.

Growth analysis revealed that during growth at 30°C, the temperature of the epithelial surface of the nasopharynx (Lindemann *et al.* 2002), *S. pneumoniae* serotypes 1, 6B, 15B, 23F, 4, and 2 showed prolonged lag phase and lower maximum growth density compared to growth at 37°C and 39°C. These findings were in contrast to Tothpal *et al.* (Tothpal *et al.* 2019) study where they found higher growth and faster lag phase in growth at 30°C compared to 37°C. However, Tothpal and colleagues obtained their growth data using the enriched brain heart infusion (BHI) media which has been previously shown to have different growth characteristics compared to CDM (Hathaway *et al.* 2012) and may explain the differences in results. Increasing deviation from 37°C has been previously associated with increased number of differentially expressed genes in *S. pneumoniae* using microarray technology (Pandya *et al.* 2005). We had hypothesized that predominantly carriage strains like serotype 6B are better adapted to the conditions of the nasopharyngeal environment and would have a different growth kinetics compared to strains with higher invasive potential like serotype 1 and serotype 4 at 30°C. Serotype 6B (BHN418) exhibited a longer lag and a slower growth rate than the more invasive strains of serotypes 1 (BVJ11JL) and serotype 4 (TIGR4). We speculate that serotype 6B, which is a frequent carriage but infrequent invasive strain, may be better adapted to growing at nasopharyngeal temperature (30°C) and that the growth reduction may be an adaptive

mechanism to ensure prolonged colonisation. But at 37°C and 39°C all strains are equally adapted. Other possible explanations for temperature differences in growth phenotype could include capsule differences between pneumococcal strains (Hathaway *et al.* 2012; Weinberger *et al.* 2009). *S. pneumoniae* regulates capsules thickness in response to environment as a key adaptive mechanism during both colonization and invasive disease (Weiser, Ferreira, and Paton 2018). Serotype 6B for example has a capsule size three times thicker than serotype 4 (Weinberger *et al.* 2009). The temperature-dependent differences observed in this chapter will be further investigated at the transcriptomic level in this thesis.

*S. pneumoniae* biosynthesis of carbohydrates is critical for capsule production (Troxler *et al.* 2019; Weinberger *et al.* 2009). Growth of our strains in disaccharide fructose resulted in the lowest maximum growth density of *S. pneumoniae* strains compared to growth in carbohydrates N-acetylglucosamine, galactose, and sucrose. Capsule thickness of *S. pneumoniae* has been previously observed to reduce when grown in fructose with more heavily encapsulated strains such as serotypes 6B being more affected (Weinberger *et al.* 2009). Differences in the intracellular concentrations of phosphorylated metabolites such as UDP-glucose and UDP-galactose which are important precursor molecules for capsule production have been implicated in phenotypic growth differences of *S. pneumoniae* in fructose (Troxler *et al.* 2019). Growth of non-encapsulated pneumococcal strains have a drastic decrease in levels of UDP-glucose and UDP-galactose when grown in fructose suggesting that *S. pneumoniae* is not able to produce critical precursors for capsule biosynthesis under fructose conditions (Troxler *et al.* 2019).

To induce metal stress, the chelator 2,2 bipyridyl (BIP) was added to growth medium. Increasing concentrations of BIP in growth media resulted in reduction in maximum growth densities of *S. pneumoniae* strains in a dose-dependent manner. These results were confirmed in enriched media THY which contains up to four-fold the levels of iron than in CDM (Troxler *et al.* 2019), suggesting an inherent adaptive mechanism by *S. pneumoniae* to environmental changes in transition metals concentration. Understanding of the mechanisms of iron regulation and acquisition in *S. pneumoniae* is still limited and no known study has evaluated the transcriptomic response of *S. pneumoniae* to iron stress

using RNA-Seq. Proteomics and morphological analysis have shown substantial impact of iron limitation on *S. pneumoniae* adaptation by affecting lipid metabolism, peptidoglycan biosynthesis, and cell division pathways (Hoyer *et al.* 2018). Pneumococcal iron restriction has also been shown to be marked by increase expression of known virulence factors such as *psaA* (Nanduri *et al.* 2008). In addition, the expression levels of a pneumococcal gene operon (SP0727 – SP0729 in TIGR4) essential for colonisation of the mouse nasopharynx was found to be sensitive to iron stress (van Opijnen and Camilli 2012). The effects of iron restriction on *S. pneumoniae* transcriptome will be compared for *S. pneumoniae* serotype 1 and serotype 6B strains as part of this thesis.

Careful planning of transcriptomic experiments is essential for obtaining robust datasets. The work in this chapter highlights the various optimisation experiments that were undertaken to select a suitable growth media, characterise growth characteristics of *S. pneumoniae* strains under different environments, and refine RNA isolation and experimental approaches for planned RNA-Seq experiments. A key consideration was the identification of a CDM that supports the growth of *S. pneumoniae* strains and enabled altering of constituents of the media. For these reasons, the NMC growth medium described by Aprianto *et al* (Aprianto *et al.* 2018) was selected as the most suitable for transcriptomic experiments in this thesis. The use of this CDM has the advantage of enabling direct comparison of our transcriptomic data from clinical strains in this thesis to the transcriptomics data of laboratory adapted D39 (serotype 2) strain (Aprianto *et al.* 2018).

Chapter 4 : Generating high-quality  
complete genome assemblies of  
*Streptococcus pneumoniae*  
serotype 1 (BVJ1JL) and serotype  
6B (BHN418) strains

## 4.1 Introduction

The first complete genome assembly of *Streptococcus pneumoniae* was generated for the clinical strain TIGR4 (serotype 4) and published in 2001 (Tettelin *et al.* 2001). The publication of the complete sequence of *S. pneumoniae* provided surprising insights such as the unexpectedly large number of sugar substrate utilisation systems in *S. pneumoniae* compared to other commensals of the human upper respiratory tract like *Haemophilus influenza* and *Neisseria meningitidis* (Buckwalter and King 2012; Tettelin *et al.* 2001). The genomic revolution has resulted in an explosion of genome sequences of bacterial strains (van Vliet 2010; Mardis 2017). For example, there are currently more than 300,000 sequenced *Salmonella* genomes (Zhou *et al.* 2020). However, most of the publicly available bacterial assemblies are fragmented draft sequences (Wick *et al.* 2017).

A primary objective in this thesis was to use RNA-Seq to investigate the transcriptomic adaptation of clinical strains of *S. pneumoniae* in different environmental conditions. We have proposed to use serotype 1 (BVJ1JL) and serotype 6B (BHN418) for comparative transcriptomic experiments in this thesis and the rationale for the selection of these strains has been outlined in Chapter 2. Functional genomics studies require high-quality genome assemblies that can be used to accurately map and quantify transcripts. Such reference quality genome assemblies were unavailable for the clinical pneumococcal strains of serotype 6B (BHN418) and serotype 1 (BVJ1JL) used in this thesis. Most of the available microbial genomes have been generated with shotgun sequencing which poorly resolves repetitive regions and results in low-quality and inaccurate genome assemblies (Koren *et al.* 2013). The best annotated pneumococcal genome assembly is for the highly laboratory adapted D39 (serotype 2) strain (Slager, Aprianto, and Veening 2018) which was generated through a hybrid assembly method. Currently, 9031 genomes of *S. pneumoniae* have been sequenced and deposited in NCBI (in March 2022) (Jolley, Bray, and Maiden 2018). By the start of this project (early 2019), forty *S. pneumoniae* assemblies were complete but only one of these was a serotype 1 genome assembly (strain INV104) which was isolated from the United Kingdom in 2010 (Hiller *et al.* 2007), with the remaining being fragmented draft sequences. The publicly available sequence of the serotype 6B strain (BHN418) being used for transcriptomic experiments in this thesis was also incomplete and made up of 68 contigs (Browall *et al.* 2014).

The aim in this chapter was therefore to use a hybrid genome assembly approach (Wick *et al.* 2017), combining PacBio®'s single molecule real-time (SMRT) long-read sequencing technology (Eid *et al.* 2009) and Illumina®'s second-generation short-read sequencing (Bentley *et al.* 2008) to generate complete pneumococcal genome assemblies of serotype 1 (BVJ1JL) and serotype 6B (BHN418) strains. SMRT technology excels in sequencing through repetitive elements and plasmids, is not sequence-context biased, and also provides epigenetic information, which adds value to final genome annotation. Illumina® short-read sequencing is notable for its base call accuracy, which is the SMRT technology weakest point. This hybrid approach of combining Illumina and PacBio sequences has proven successful in generating high-quality complete genome assemblies of bacterial pathogens (Kamada *et al.* 2014; Garzetti *et al.* 2017; Bashir *et al.* 2012). Crucially, this approach often leads to the identification of novel genes and previously unidentified functional elements in bacterial genomes (Slager, Aprianto, and Veening 2018; Bashir *et al.* 2012; Kamada *et al.* 2014).

The generated reference-quality genome assemblies of serotype 1 (BVJ1JL) and serotype 6B (BHN418) will be used for efficient RNA-Seq transcripts mapping and quantification later in this thesis.

## 4.2 Methods

### 4.2.1 Bacterial strains

The serotype 1 strain (BVJ1JL) was isolated from the nasopharynx of a 9-year child in Blantyre, Malawi in 2015 as part of a community surveillance of *S. pneumoniae* prevalence (Swarthout, Fronterre, *et al.* 2020). The serotype 6B strain (BHN418) was isolated in Norway between 1997 – 2004 (Browall *et al.* 2014). The BHN418 strain is also the strain that is being used to inoculate healthy adults as part of the controlled experimental human challenge model at Liverpool (Wright *et al.* 2012). The draft genome sequence of BHN418 has previously been generated using short read sequencing and deposited in GenBank (Browall *et al.* 2014). Thus, we have referred to the re-sequence genome assembly of BHN418 in this thesis as BHN418UCL.

### 4.2.2 Microbiology culture of *S. pneumoniae*

The BHN418 strain was obtained as a purified single colony stored in glycerol at -80°C from the laboratory of Prof. Jeremy Brown (University College London). The stock used in the present study is the same as a previous study from the Heyderman laboratory (UCL) assessing pneumococcal interaction with primary epithelial cell lines (Weight *et al.* 2018). The serotype 1 (BVJ1JL) strain was obtained as a nasopharyngeal sample (NPS) stored in STGG at -80°C. The NPS sample of BVJ1JL was previously shown by microarray analyses to be a single carriage serotype 1 sample (Swarthout, Gori, *et al.* 2020). Microbiological identification of BVJ1JL as a serotype 1 strain was assessed again in this thesis by morphology, optochin test, bile solubility test, and Gram stain. Capsular type of the BVJ1JL was confirmed as serotype 1 by serological latex agglutination test kit (ImmuLex™ 7-10-13-valent Pneumotest; Statens Serum Institute, Denmark) (Chapter 2) and later by whole genome sequencing (PneumoCaT). All cultures were performed using Columbia blood agar (CBA) supplemented with horse blood and incubated at 37°C and 5% CO<sub>2</sub> for ~18hrs.

### 4.2.3 Genomic DNA isolation

Unlike short read sequencing platforms like Illumina, PacBio does not rely on PCR amplification cycles and directly sequences input DNA to produce contiguous sequencing

reads often in excess of 10 kilobases (Eid *et al.* 2009). Thus, specialised library preparation methods are required for PacBio sequencing beginning with high molecular weight DNA (Mardis 2017, Eid *et al.* 2009). The QIAGEN genomic-tip 500/G DNA kit was used for isolation of high molecular weight DNA. To enhance lysis of pneumococcal capsule, the Qiagen lysis buffer was supplemented with 100mg/mL of lysozyme (Sigma Aldrich) and 50U mutanolysin (Sigma Aldrich). Fresh overnight pneumococcal cultures in CBA incubated at 37°C with 5% CO<sub>2</sub> were harvested into lysis buffer supplemented with 100mg/mL of lysozyme (Sigma Aldrich) and 50U mutanolysin (Sigma Aldrich). Proteinase K (Qiagen) and RNase A (100mg/mL) (Qiagen) were used as part of manufacturer's protocol to degrade cellular proteins and RNA respectively. Isolation of genomic DNA was then performed according to manufacturer's instructions. DNA for Illumina short-read sequencing was isolated using DNeasy blood and tissue kit (Qiagen) following manufacturer's instructions. The lysis buffer of the DNeasy kit was also supplemented with 100mg/mL of lysozyme (Sigma Aldrich) and 50U mutanolysin (Sigma Aldrich). Quality and quantity of isolated DNA was assessed using nanodrop spectrophotometer (Thermo Fisher) and Qubit™ (Invitrogen) (using DNA broad range (BR) Assay kit).

#### **4.2.4 PacBio long read sequencing**

PacBio sequencing was performed for *S. pneumoniae* serotype 6B (BHN418) and serotype 1 (BVJ1JL) at the University of Exeter Sequencing facility. High molecular weight genomic DNA was sheared using g-TUBE (Covaris) with a target length of 10Kb. The DNA was then profiled on an Agilent Tapestation 2200 using genomic tapes. Library preparation was performed according to the protocol "Preparing Multiplexed Microbial Libraries Using SMRTbell® Express Template Prep Kit 2.0" (Pacific Biosciences) with barcoded overhanging adaptors 8A (Pacific Biosciences). The samples were then pooled, purified and size selected >3Kb with AMPure PB beads (Pacific Biosciences) as recommended in manufacturer protocol. Qubit fluorometry (Thermofisher) and Genomic tape profiling (Agilent) were then performed to allow annealing and polymerase binding of the SMRTbell library template using Sequencing primer v4 and Sequel binding kit v3 (Pacific Biosciences). A 10h capture using an insert size of 4Kb in a single 1M SMRT cell was then performed on the PacBio Sequel using Sequel v3 chemistry.



#### 4.2.5 Illumina short read sequencing

The short reads sequences for BVJ1JL (serotype 1) were previously generated at Oxford Genomics Centre using Illumina High Seq 4000 platform with read length of 150bp (Swarthout, Gori, *et al.* 2020). DNA was quantified using Quant-iT PicoGreen double-stranded DNA (dsDNA) kits (Invitrogen) according to the manufacturer's specifications and fragmented using an EpiSonic sonication system (EpiGentek). The libraries were constructed using the NEBNext Ultra DNA sample prep master mix kit (New England BioLabs) with in-house adapters and barcode tags described at Oxford Genomics Centre, UK (Lamble *et al.* 2013). The libraries were sequenced on an Illumina HiSeq 4000 instrument as 150bp paired reads. Paired-end short read sequences for BHN418 were generated at UCL Pathogens Genomic Unit using the Illumina MiSeq platform with read length of 150bp.

#### 4.2.6 Sequence reads quality control

Initial processing of PacBio sequencing output such as demultiplexing and down sampling was performed using SMRTLink portal v8.0 at the University of Exeter Sequencing facility. Sequencing metrics for PacBio reads were obtained using NanoPlot v1.38.0 (De Coster *et al.* 2018). Raw Illumina reads were first assessed with FastQC (v0.11.5) (S. 2010) and low-quality and adapter sequence reads were removed using Trimmomatic v0.39 (Bolger, Lohse, and Usadel 2014) using minimum average phred quality of 20 within any given 4 nucleotides sliding window. Trimmed reads were again checked and visualised by FastQC v0.11.5, Babraham Bioinformatics, UK to confirm removal of low-quality reads (S. 2010). Seqtk was used to sub sample 700,000 illumina reads to be used for hybrid genome analysis.

#### 4.2.7 Genome assembly and quality assessment

*de novo* assembly of PacBio long reads only was performed with Hierarchical Genome Assembly Process (HGAP4) within the SMRTLink portal at the University of Exeter. Hybrid genome assembly combining the Illumina short reads and PacBio long reads was performed using Unicycler (v0.4.9b) (Wick *et al.* 2017). The generated assemblies were

visualised using Bandage (v0.8.1) to confirmed number of contigs and check for genome circularisation (Wick *et al.* 2017). Unicycler rotates bacterial genomes and implement genome circularization. Quast v5.1.0rc1 was used to assess quality of assemblies and to generate assembly metrics (Gurevich *et al.* 2013). The D39V strain was used as a reference for Quast analysis (Slager, Aprianto, and Veening 2018).

#### **4.2.8 Genome annotation and curation**

Automatic annotation of the BVJ1JL and BHN418UCL assemblies were performed using prokka (v1.14.6) (Seemann 2014). Prokka pipeline was Implemented to be GenBank complaint and was run with `–rfam` command to utilise Rfam to annotate non-coding RNA listed on the Rfam database (Kalvari *et al.* 2018). The integrity of the Prokka annotation was further assessed using the [NCBI website](#). Genome integrity checks included taking neighbouring pairs of proteins to do BLASTP analysis. When a pair does hit the same longer protein, it suggests that the pair may represent a single gene that has gained frameshift or other mutations along the way. Manual analysis of such pairs was conducted by BLAST to decide whether the pair should remain two proteins or be combined into a single pseudogene. Manual curation was conducted using [NCBI Genome Workbench](#) v3.6.0 using the MacBook pro desktop application.

#### **4.2.9 Genetic characterisation of BVJ1JL (serotype 1) and BHN418UCL (serotype 6B)**

Genetic characterisation of the sequenced genomes was done using various publicly available software's. To confirm strain and MLST types of the strains the mlst program was used. The serotype of the strains was confirmed *in silico* using SeroBa and PneumoCaT. Abricate and ResFinder were used to query genome assemblies against a database of known resistance genes. Plasmid finder and Phaster were used to identify potential presence of plasmids and prophages in the assembled genome sequences.

#### **4.2.10 Comparative genomics and pangenome analysis**

Roary (Page *et al.* 2015) was used to identify pangenome and core genome of BHN418UCL (serotype 6B) and BVJ1JL (serotype 1) together with a list of relevant reference quality genomes (**Table 4.1**). This was important to identify orthologs between clinical strains of

BVJ1JL (serotype 1) and BHN418UCL (serotype 6B) and other reference strains for subsequent comparative transcriptomics analyses. A brief history of all strains used and the rationale for selection is presented in **Table 4.2**. Visualisation of genomic differences in the assemblies of BHN418, BHN418UCL, and the D39V (serotype 2) reference genome assembly was done using the BRIC software. The assembly and annotation of D39V was used because it is the best annotated pneumococcal genome sequence which was also assembled using hybrid methods (Slager, Aprianto, and Veening 2018). Roary generated core genome results were visualise using custom scripts.

**Table 4.1:** Characteristics of complete *S. pneumoniae* genome sequences used for comparative genomics and identification of orthologous genes

	TIGR4	D39V	BHN418	INV104	670-6B	BVJ1JL	BHN418UCL
Serotype	4	2	6B	1	6B	1	6B
# Contigs	1	1	68	1	1	1	1
# Bases	2,160,837	2,046,572	2,063,372	2,142,122	2,240,045	2,134,668	2,106,999
# Genes	2187	2077	2165	1941	2352	2,238	2228
Source	disease	disease	carriage	disease	disease	carriage	carriage
Year	1995	1916	1998	2010	1988	2015	1998
MLST	205	ST595	ST138	ST227	ST90	ST5012	ST138
Assembly	complete	complete	draft	complete	complete	complete	complete
Sequence Method	Sanger	PacBio + Illumina	454, Sanger	Sanger	Sanger	PacBio + Illumina	PacBio + Illumina
Country	Norway	USA	Sweden	UK	Spain	Malawi	Sweden
Reference	(Tettelin <i>et al.</i> 2001)	(Slager, Aprianto, and Veening 2018)	(Browall <i>et al.</i> 2014)	(Hiller <i>et al.</i> 2007)	(Donati <i>et al.</i> 2010)	(Betts <i>et al.</i> 2021)	This thesis

**Table 4.2:** *S. pneumoniae* reference genome assemblies downloaded from RefSeq to identify pneumococcal orthologs. The rationale for selecting these reference genomes is stated.

Strain	Serotype	Rationale	Accession No.	Ref
TIGR4	4	Commonly used strain in <i>S. pneumoniae</i> research. TIGR4 was the first complete genome sequence of <i>S. pneumoniae</i> and has been recently reannotated.	NC_003028	(Tettelin <i>et al.</i> 2001)
D39V	2	Arguably the most common <i>S. pneumoniae</i> strain used in research. D39V has been recently re-sequenced and assembled.	NZ_CP027540.1	(Slager, Aprianto, and Veening 2018)
670-6B	6B	One of the few <i>S. pneumoniae</i> strains that has annotated pathway information in KEGG database. Orthologs of this strain will be used to identify pathways in BHN418UCL.	CP002176.1	(Donati <i>et al.</i> 2010)
INV104	1	One of the few <i>S. pneumoniae</i> strains that has annotated pathway information in KEGG database. Orthologs of this strain will be used to identify pathways in BVJ1JL.	FQ312030.1	(Hiller <i>et al.</i> 2007)
BHN418	6B	The original genome sequence of BHN418 strain generated by shotgun sequencing technology. Comparison to BHN418UCL will enable identification of differences in genome assemblies.	ASHP000000000.1	(Browall <i>et al.</i> 2014)
BVJ1JL	1	The BVJ1JL strain described in this project has been published and included as reference strain here. The locus tags of this strain will be used for genome comparison analysis.	CP071871.1	(Betts <i>et al.</i> 2021)

### 4.3 Results

#### 4.3.1 Genomic DNA quantity and quality

High molecular weight *S. pneumoniae* genomic DNA was successfully isolated using the QIAGEN genomic-tip 500/G DNA kit with modifications to the manufacturer's protocol. Manufacturer's protocol was modified by supplementing the lysis buffer with 100mg/mL of lysozyme (Sigma Aldrich) and 50U mutanolysin (Sigma Aldrich). Lysozyme and mutanolysin were added to enhance chemical breakdown of the thick *S. pneumoniae* polysaccharide capsule. Mechanical breakdown of cell wall by bead beating would shear DNA and therefore an inappropriate approach to obtain high molecular weight DNA. Molecular weight and purity of isolated DNA was assessed by gel electrophoresis and DNA was confirmed to be of pure quality with nanodrop values for 280/260 ratio of ~1.8 for both strains (**Table 4.3**). These quantification values were adequate for PacBio long read sequencing.

**Table 4.3:** Quantification results of genomic DNA isolated from *S. pneumoniae* strains BVJ1JL (serotype 1) and BHN418 (serotype 6B) for PacBio sequencing

Strain	Qubit (BR DNA) ng/μL	Nanodrop		
		ng/μL	280/260	260/230
Serotype 6B (BHN418UCL)	460	435.5	1.91	1.42
Serotype 1 (BVJ1JL)	604	661.9	1.81	1.66

#### 4.3.2 PacBio reads sequencing metrics

The genomic DNA of BVJ1JL (serotype 1) and BHN418UCL (serotype 6B) were sequenced in PacBio Sequel to genome coverage depths of ~698x and ~2505x respectively (**Table 4.5**). BHN418UCL was sequenced at ~4x higher coverage than BVJ1JL because the re-quantification of the genomic DNA at the sequencing facility indicated that the recovery of DNA was lower for BHN418UCL compared to BVJ1JL after shearing (necessary for barcoding) suggesting potential degradation. The higher sequencing depth of BHN418UCL

was therefore done to compensate for potential effects of DNA degradation. However, due to the extremely high number of resulting reads for BHN418UCL, *de novo* assembly in HGAP4 was challenging with the entire reads generated. After subsampling down the reads to 25 % of total reads for BHN418UCL, *de novo* assembly of PacBio long reads was achieved in SMRT link. The subsampled reads still represented >500x coverage of the BHN418UCL genome. The sequencing outputs generated for both BHN418UCL (serotype 6B) and BVJ1JL (serotype 1) using the PacBio sequencer are summarised in **Table 4.4**. Sequencing metrics of PacBio sequencing reads as estimated by NanoPlot (De Coster *et al.* 2018) are presented in **Table 4.5**.

**Figure 4.4:** Raw sequencing output for BVJ1JL (serotype 1) and BHN418UCL (serotype 6B) *S. pneumoniae* strains sequenced with PacBio Sequel.

Sample	Barcode	Polymerase reads	Subreads	Bases	Mean Length	Mean Longest Subread
BVJ1JL	bc1001_BAK8A	49,945	325,793	1,466,877,255	29,812	6,280
BHN418UCL	bc1002_BAK8A	182,255	1,011,284	5,261,915,926	29,236	7,187

**Table 4.5:** Summary statistics of PacBio fastq reads estimated by NanoPlot (De Coster *et al.* 2018)

Metric	Serotype 1 (BVJ1JL)	Serotype 6B (BHN418UCL)
<b>General Summary</b>		
Mean read length	4,918.9	5,740.5
Mean read quality	31.8	30.4
Median read length	4,455.0	5,292.0
Median read quality	29.5	28.3
Number of reads	16,057.0	48,015.0
Read length N50	5,131.0	6,218.0
STDEV read length	1,954.4	2,252.9
Total bases	78,983,343.0	275,629,152.0
<b>Number, percentage and megabases of reads above quality cutoffs</b>		
>Q5	16057 (100.0%) 79.0Mb	48015 (100.0%) 275.6Mb
>Q7	16057 (100.0%) 79.0Mb	48015 (100.0%) 275.6Mb
>Q10	16057 (100.0%) 79.0Mb	48015 (100.0%) 275.6Mb
>Q12	16057 (100.0%) 79.0Mb	48015 (100.0%) 275.6Mb
>Q15	16057 (100.0%) 79.0Mb	48015 (100.0%) 275.6Mb
<b>Top 5 highest mean basecall quality scores and their read lengths</b>		
1	93.0 (1864)	93.0 (1972)
2	93.0 (1679)	93.0 (1457)
3	93.0 (1046)	93.0 (328)
4	93.0 (389)	93.0 (872)
5	93.0 (148)	93.0 (2233)
<b>Top 5 longest reads and their mean basecall quality score</b>		
1	19136 (20.8)	20609 (22.1)
2	17767 (22.0)	20361 (20.7)
3	17247 (21.0)	20045 (20.5)
4	17219 (20.8)	19368 (21.6)
5	16861 (20.4)	19294 (21.5)



### 4.3.3 Illumina sequencing reads metrics

The total number of raw Illumina DNA reads for serotype 1 (BVJ1JL) was 1,162,988 paired end reads. Following removal of low-quality reads and adapters, a total 700,000 paired end reads (1,400,000 reads in total) were used for hybrid genome assembly. This represented genome coverage of approximately 50x. Total number of raw short-read Illumina sequences for serotype 6B (BHN418UCL) was 2,951,794 paired end reads. Following quality control trimming, 2,395,233 paired-end Illumina short-read sequences were used for genome assembly of the BHN418UCL genome.

### 4.3.4 HGAP *de novo* assembly metrics

Microbial assembly protocol with HGAP within [SMRT Link](#) (v8.0.080529) was used for *de novo* assembly of PacBio sequences only. These assembly metrics are summarised in **Table 4.6**. *de novo* assembly with HGAP produced two contigs for each strain with one of the contigs for serotype 1 (BVJ1JL) appearing to be a complete chromosome coverage of *S. pneumoniae* with genome length of 2,154,369bp. However, none of the contigs were circularised in the resulting output.

**Table 4.6:** Characteristics of *de novo* assembly of PacBio sequenced with HGAP within SMRTLink (v8.0)

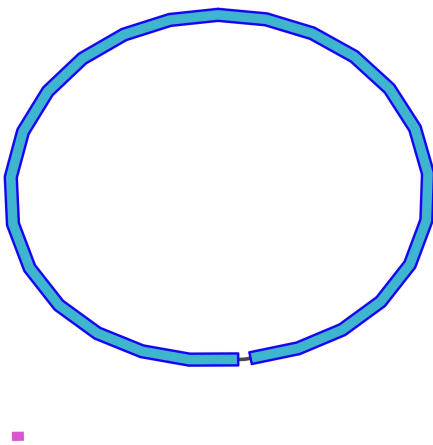
Metric	BHN418UCL (serotype 6B)	BVJ1JL (serotype 1)
Polished Contigs	2	2
Length of contig 1	1,688,502	2,154,369
Length of contig 2	434,041	12,337
N50 Contig Length	1,688,502	2,154,369
Sum of Contig Lengths	2,122,543	2,166,706
Mean Coverage	~517x	~573x
Percent Aligned Bases	87.66%	88.41%
Number of Subread Bases (aligned)	1,144,351,142	1,293,329,307
Number of Subreads (aligned)	227,836	296,798
Number of Alignments	229,734	300,038
Alignment Length Mean (aligned)	4,981	4,311
Alignment Length N50 (aligned)	5,913	4,883
Alignment Length 95% (aligned)	9,840	8,431
Alignment Length Max (aligned)	25,208	22,889
Number of Polymerase Reads (aligned)	41,725	46,138
Polymerase Read Length Mean (aligned)	28,844	29,418
Polymerase Read N50 (aligned)	51,295	52,042
Polymerase Read Length 95% (aligned)	83,826	85,190
Polymerase Read Length Max (aligned)	125,978	125,843

#### 4.3.5 Hybrid genome assembly of BHN418UCL (serotype 6B) and BVJ1JL (serotype 1) *S. pneumoniae*

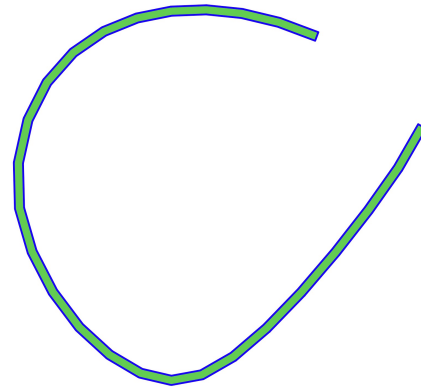
Unicycler was used to combine the PacBio, and Illumina reads to construct genome assemblies (Wick *et al.* 2017). Hybrid *de novo* assembly of serotype 1 (BVJ1JL) reads produced two contigs, a full-length circularised chromosome of length 2.1Mb and a smaller linear contig of length ~1.2Kb (**Figure 4.1A**). BLAST and alignment analysis of the smaller 1.2Kb sequence in BVJ1JL was found to be identical (99.59%) to a part of the larger

contig and containing a highly repetitive sequence. The smaller contig was therefore discarded and the larger circularized contig considered the complete genome sequence of the serotype 1 (BVJ1JL) strain. Unicycler hybrid assembly resolved the reads of the serotype 6B (BHN418UCL) strain into a single linear contig of length 2,106,999bp (**Figure 4.1B**). These full-length hybrid genome assemblies were used for all subsequent annotation and genetic characterisations of the serotype 6B (BHN418UCL) and serotype 1 (BVJ1JL) strains.

### A). Serotype 1 (BVJ1JL)



### B). Serotype 6B (BHN418UCL)



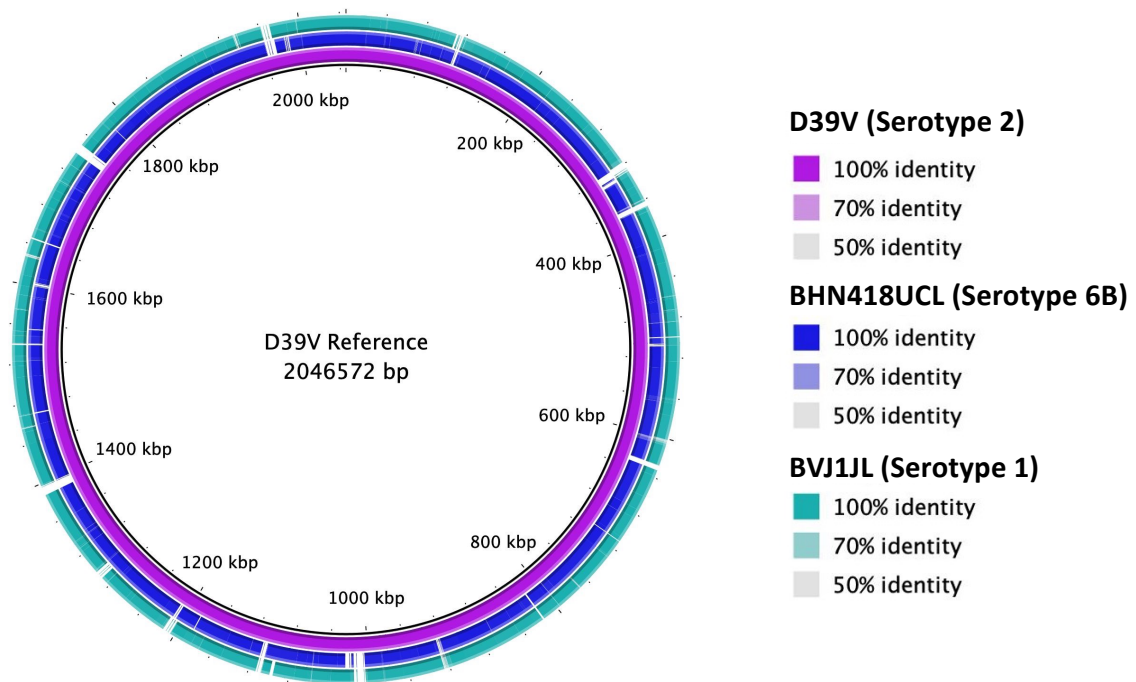
**Figure 4.1:** Visualisation of *S. pneumoniae* serotype 1 (BVJ1JL) and serotype 6B (BHN418UCL) genomes assembled by combining long-read and short-read sequencing reads using Unicycler. **A).** Unicycler resolved the BVJ1JL genome into a single circularised full-length chromosome of size 2,134,668bp (blue) and a second contig of size ~1.2Kb in length (pink). The 1.2Kb contig was later found to form part of a highly repetitive region of the larger chromosome and removed from assembly. **b** The BHN418UCL genome was resolved into a single linear chromosome of length 2,106,999bp (green).

### 4.3.6 Genome annotation and pangenome analysis

Genome assemblies annotated by prokka were used for pangenome analysis with roary. The annotation results of BVJ1JL (serotype 1), BHN418UCL (serotype 6B), and the D39V (serotype 2) reference sequence are presented in **Table 4.7**. The pangenome of these three genomes was 2681 genes of which 1629 were core between the strains. The core genome of serotype 1 (BVJ1JL) and serotype 6B (BHN418UCL) strains only was 1674 genes with a pangenome of 2338 genes. The genetic differences between genome assemblies of the strains of BVJ1JL (serotype 1), BHN418UCL (serotype 6B) and D39 (serotype 2) is visualised in **Figure 4.2** using D39 as reference. Pangenome analysis of the *S. pneumoniae* reference genome assemblies presented in **Table 4.1** together with BHN418UCL (serotype 6B) and BVJ1JL (serotype 1) resulted in 3033 genes in total. Of these 1473 genes were core among all the reference genomes in **Table 4.1**. These results are consistent with previous pangenome analyses of *S. pneumoniae* (Donati *et al.* 2010; Kulohoma *et al.* 2015).

**Table 4.7:** *de novo* assembly and annotation metrics of *S. pneumoniae* serotype 1 (BVJ1JL) and serotype 6B (BHN418UCL) in comparison to serotype 2 (D39) reference assembly (Slager, Aprianto, and Veening 2018).

Metric	Serotype 1 (BVJ1JL)	Serotype 6B (BHN418UCL)	Serotype 2 (D39V)
<b><i>de novo</i> assembly</b>			
Genome Coverage	>500x	>500x	250
Number of contigs	1	1	1
G+C content (%)	39.73	39.61	39.71
Contig length	2,134,668	2,122,543	2,046,572
<b>Genome annotation</b>			
Total features	2294	2243	2149
Coding sequences	2101	2076	1999
Number of genes	2238	2076	2077
Transfer RNAs	59	58	58
Ribosomal RNAs	12	12	12
Non-coding RNAs	65	81	79



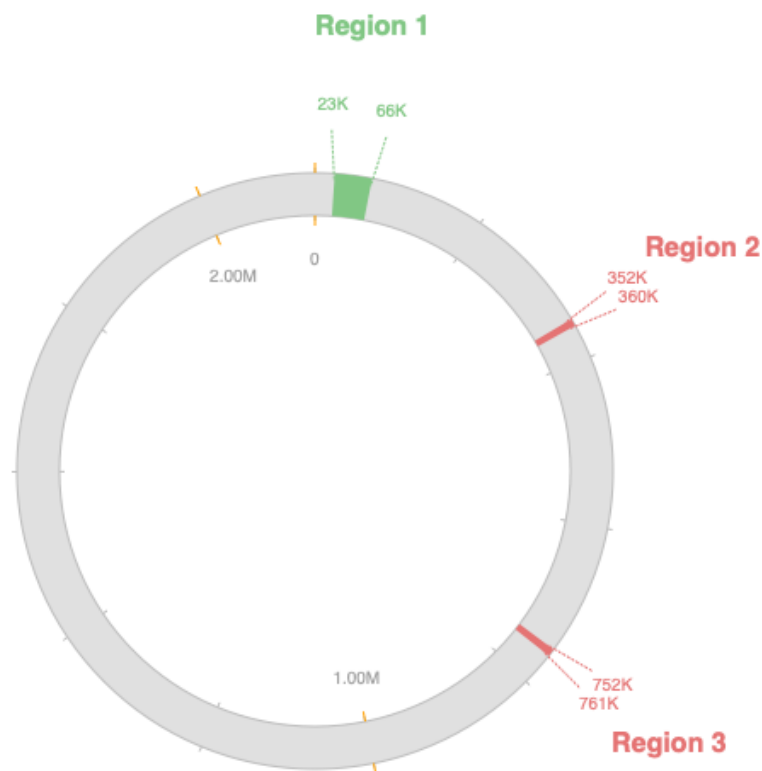
**Figure 4.2:** Circos plot of D39V (serotype 2), BHN418UCL (serotype 6B), and BVJ1JL (serotype 1) *S. pneumoniae* showing visual representation of the genetic differences between the strains. The D39V strain was used as the reference. Solid lines between genomes indicate genetic similarity between strains and gaps are indicative of absence of genomic regions. Colours correspond to the difference strains.

#### 4.3.7 Genetic characterisation of genome assemblies

The serotype of both BHN418UCL (serotype 6B) and BVJ1JL (serotype 1) were confirmed by *in silico* analysis to be serotype 6B and serotype 1 respectively. BVJ1JL belonged to sequence type 5012 (ST5012). ST5012 is a locus variant of the highly virulent ST217 (Jacques *et al.* 2020; Bricio-Moreno *et al.* 2020). Genetic analysis predicted susceptibility to penicillin but resistance to chloramphenicol and tetracycline (*cat* and *tetM*, respectively). The macrolide resistance genes *mef* and *ermB* were also not detected in BVJ1JL (serotype 1) strain.

#### 4.3.8 Prophage detection

Querying for the presence of phage in BVJ1JL (serotype 1) resulted in a total of three prophage regions. One of these prophages, of length 42.9Kb (Region 1 in **Figure 4.3**), was complete and intact (*Streptococcus* phage PH10; Accession: NC\_012756) in the BVJ1JL strain. The intact prophage was predicted to contain 69 open reading frames (proteins). The two incomplete phage regions were of length 8.4Kb (Region 2 in **Figure 4.3**; *Escherichia* phage *vB\_EcoM\_Schickermoser*; Accession: NC\_048196) and 8.8Kb (Region 3 in **Figure 4.3**; *Paenibacillus* phage *Leyra*; Accession: NC\_048692) respectively. No intact prophages were identified in BHN418UCL (serotype 6B) strain. A 11.5Kb prophage classified as questionable (score below intact) was found in the BHN418UCL (serotype 6B) genome assembly. The *Ralstonia* phage *phiRSA1* which was the best match for 11.5Kb questionable phage for BHN418UCL has a complete length of 38.7Kb.



**Figure 4.3: Prophages present in BVJ1JL (serotype 1) strain:** Genome viewer showing three prophage regions identified in BVJ1JL (serotype) 1 strain. Region 1 (green) is a complete and intact phage of *Streptococcus* phage PH10 of length 42.9Kb. Region 2 and Region 3 (red) in are incomplete phage genomes. Location of phage regions in circle corresponds to approximate position in circularised genome of BVJ1JL. Data and visualisation was generated from Phaster website.

#### 4.4 Discussion

With a goal of gaining insights into the underlying mechanisms that mediate differences associated with *S. pneumoniae* clinical strains, serotype 1 (BVJ1JL) and serotype 6B (BHN418) were selected as model strains for comparative transcriptomics experiments in this thesis. *S. pneumoniae* serotype 1 and serotype 6B exhibit contrasting propensities to colonise the host and to cause invasive disease (Brueggemann *et al.* 2004; Weinberger *et al.* 2010). However, for robust analysis of RNA-Seq datasets, a high-quality reference genome assembly is required for accurate mapping and quantification of transcripts (Croucher and Thomson 2010; Sharma and Vogel 2014). In this Chapter, genome assemblies of clinical strains of serotype 1 (BVJ1JL) and serotype 6B (BHN418UCL) were generated by combining the strength of short-read and long-read sequencing technologies (Wick *et al.* 2017; Bashir *et al.* 2012; Koren *et al.* 2013). The genome assemblies for both strains were resolved into a single contig at high genome sequencing depth (>500x). These assemblies will be used for RNA-Seq transcripts quantifications in later chapters in this thesis and will provide a valuable resource to the wider scientific community.

Despite about 10,000 *S. pneumoniae* genomes being sequenced and publicly accessible by the start of work on this thesis, less than 1% of these were classified as complete. None of the completed and publicly available *S. pneumoniae* genome assemblies were of the strains of interest for this thesis. The majority of publicly available bacterial genome sequences are generated using short-gun (e.g., Illumina) sequencing methods (van Vliet 2010; Mardis 2017). Bacterial genomes contain many repetitive sequences that are usually larger than the short fragment lengths of Illumina sequencing (500bp or less), and thus preventing assembly tools from accurately and fully resolving bacterial genomes from such reads. Reads from short-read sequencers usually result in dozens of contigs to form fragmented draft genome assemblies (Wick *et al.* 2017; van Vliet 2010; Mardis 2017). The low cost and accuracy of short-reads sequencing such as the Illumina platform makes it adequate for bacterial investigations such as genome-wide association studies. However, for robust transcriptomics analysis, where a single nucleotide polymorphism (SNP) can have significant impact on bacterial gene expression and phenotype (Hammarlof *et al.* 2018), the availability of a contiguous and fully resolved genome

assembly of the strain of interest is critical. Long-read sequencing methods such as PacBio can generate read lengths of 10Kb and longer covering most bacterial short repetitive fragments. However, Pacbio sequencers have a higher cost per base and are more prone to errors per-base than Illumina sequencing platforms (up to 15% vs 1%), although they can be sufficient to complete bacterial genomes at very high sequencing depth (Wick *et al.* 2017; Chin *et al.* 2013).

Similar to the methods applied here, Slager *et al* (Slager, Aprianto, and Veening 2018) have re-sequenced the D39 strain (serotype 2) by combining PacBio and Illumina reads resulting in a single circularised genome assembly. This D39 genome was subsequently used to investigate the transcriptomic response of D39 to various environmental conditions (Aprianto *et al.* 2018). The re-sequenced D39 genome has been claimed by the authors to be the best annotated *S. pneumoniae* genome assembly (Slager, Aprianto, and Veening 2018). Pangenome analysis of assembled genomes identified the core genome of BHN418UCL (serotype 6B) and BVJ1JL (serotype 1) to consist of 1674 genes, consistent with previous reports (Donati *et al.* 2010; Kulohoma *et al.* 2015). Given the high number of orthologs between serotype 1 and serotype 6B, we hypothesise that the differences in the behaviour of serotype 1 and serotype 6B as shown in growth analysis in Chapter 3 will be associated with differences in the regulation of orthologous genes between the strains. This hypothesis will be tested in later chapters when the transcriptomics adaptation of the serotype 1 and serotype 6B strains are investigated under specific environmental conditions.

The generated assembly of serotype 1 (BVJ1JL) was found to contain an intact *Streptococcus* PH10 prophage. PH10 prophage is a temperate phage which has been previously identified in related *S. pneumoniae* species of *S. mitis* and *S. oralis* (van der Ploeg 2010). About half of clinical *S. pneumoniae* strains harbour prophages suggesting potential contribution to *S. pneumoniae* fitness and virulence (Romero, Garcia, and Mitchell 2009). *in silico* genetic characterisation of the assembled genomes of *S. pneumoniae* BVJ1JL (serotype 1) and BHN418UCL (serotype 6B) confirmed the serotype of the strains, presence of antimicrobial genes, and genetic relatedness between the strains.



In conclusion, the work in this chapter achieved the aim of generating a high-quality reference genome sequences of *S. pneumoniae* serotype 1 (BVJ1JL) and serotype 6B (BHN418UCL) which will be critical for robust comparative transcriptomics analyses in later chapters in this thesis.

Chapter 5 : Assessing the stress  
response of *S. pneumoniae*  
mutants lacking the biosynthesis  
genes *fhs* and *proABC*

## 5.1 Introduction

Several murine and human studies have shown that *S. pneumoniae* colonisation has an immunising effect in the host mucosa by eliciting immunological responses that protect against recolonisation and invasive disease (Ferreira *et al.* 2013; Roche, King, and Weiser 2007; Jochems *et al.* 2019; Zhang, Clarke, and Weiser 2009; Granat *et al.* 2009). Therefore, deliberate colonisation of the nasopharynx with harmless *S. pneumoniae* strains may have the potential to cross-protect against invasive pneumococcal strains, providing a viable preventive strategy to build natural immunity (Ramos-Sevillano, Ercoli, Guerra-Assunção, *et al.* 2021; Ramos-Sevillano, Ercoli, Felgner, *et al.* 2021). This strategy could provide broader host protection against pneumococcal diseases compared to polysaccharide conjugate vaccines whose efficacy is limited to pneumococcal serotypes that they contain (Feldman and Anderson 2020). Advances in the development of interventions that offer broader serotype independent protection could considerably reduce the burden of pneumococcal disease.

As potential live *S. pneumoniae* strains that are unable to cause disease but still elicit a robust host adaptive immune response, Ramos-Sevillano *et al.* (Ramos-Sevillano, Ercoli, Felgner, *et al.* 2021) have created *S. pneumoniae* mutant strains in a serotype 6B background (strain BHN418), one with a deletion of biosynthesis genes formate-tetrahydrofolate ligase ( $\Delta fhs$ ) and the other with glutamate 5-kinase/gamma-glutamyl phosphate reductase/pyrroline-5-carboxylate reductase ( $\Delta proABC$ ) deletion. Both colonise the nasopharynx as effectively as wild type (WT) strains but are attenuated for virulence in murine models. To minimise risk of the mutants reverting to WT phenotypes *in vivo*, the virulence iron transporter gene *piaA*, was also deleted to generate double mutants ( $\Delta fhs/piaA$  and  $\Delta proABC/piaA$ ). Colonisation of mice with these two mutants elicited immunological responses that were also cross-protective against *S. pneumoniae* serotype 2 (strain D39) and serotype 4 (strain TIGR4) (Ramos-Sevillano, Ercoli, Felgner, *et al.* 2021). Expanding on these findings, unpublished data by Dr Caroline Weight from the Heyderman laboratory at UCL, using previously published approaches (Weight *et al.* 2018), showed that the double mutants differentially associate with and invade epithelial cells *in vitro* and that the  $\Delta fhs/piaA$  mutant activates unique epithelial transcriptomic profiles compared to wild type strains.

Given the attenuation of the mutant strains for virulence, the aim is to test these as potential vaccine platforms in a pneumococcal controlled human infection model in Liverpool (Gritzfeld *et al.* 2013). However, the biological mechanisms that link deletion of *fhs* and *proABC* genes to attenuation of virulence *in vivo* and *in vitro* require further characterisation. *fhs* encodes a metabolic enzyme that catalyses the formation of 10-formyl-tetrahydrofolate for purine biosynthesis and is also involved in the biosynthesis and transport of methionine, serine, glycine, and thymine (Afzal, Shafeeq, and Kuipers 2016; Crowley *et al.* 1997). The *fhs* gene has been suggested as a potentially essential bacterial protein in two large-scale high-throughput screening studies for identification of *S. pneumoniae* virulence determinants undertaken by transposon-insertion sequencing (Tn-Seq) (van Opijnen and Camilli 2012) and CRISPR inference sequencing (CRISPRi-seq) (Liu *et al.* 2021). Furthermore, an *S. pneumoniae fhs* gene knockout mutant was avirulent in a murine model of invasive disease and had reduced adherence to epithelial cells compared to WT strains (Mahdi *et al.* 2012). The *proABC* proteins are a sub pathway of the proline biosynthesis pathway (and by extension the amino-acid biosynthesis pathway) that synthesises L-proline from L-glutamate (UniProt 2021; Kanehisa 2000; Belitsky *et al.* 2001). In murine models of sepsis and pneumonia, deletion of *proB* attenuates pneumococcal mortality, and delayed lethality in influenza A virus (IAV) infected mice by 40% compared to WT (Smith *et al.* 2021).  $\Delta$ *proB* *S. pneumoniae* strains have also been shown by genomic array footprint (GAF) screening to be attenuated in an experimental meningitidis study suggesting it may be essential for *S. pneumoniae* replication during meningitis (Molzen *et al.* 2011). Like *fhs*, very little is known about the functional mechanism of *proABC* and its role in *S. pneumoniae* pathogenesis.

We hypothesise that beyond the individual genes themselves, deletion of the *fhs* or *proABC* genes leads to direct or indirect dysregulation of a range of *S. pneumoniae* virulence genes leading to an attenuated *in vivo* and *in vitro* phenotype. To gain biological insights into functional properties of the mutants, we used a mutant RNA-Seq dataset generated by Dr Elisa Ramos-Sevillano in the laboratory of Prof. Jeremy Brown at UCL to compare the global gene expression of  $\Delta$ *fhs* and  $\Delta$ *proABC* *S. pneumoniae* relative to the WT strains in enriched media (THY) and under serum stress.

The aim of this chapter was to develop and test a *S. pneumoniae* whole transcriptome analysis pipeline for identification of differentially expressed genes, enrich pathways, and biological modules. The pipeline was incorporated in analyses investigating pneumococcal transcriptomic response to changes in iron concentrations and temperature detailed in Chapter 6 and Chapter 7 respectively.

## 5.2 Methods

The initial processing was conducted by Dr Elisa Sevillano Ramos (sections 5.2.2 to 5.2.4). I performed the subsequent processing and analysis (section 5.2.5 onwards).

### 5.2.1 Mutant genes information

The *fhs* and *proABC* genes are highly conserved in *S. pneumoniae* strains (Liu *et al.* 2021).

**Table 5.1** describes the *fhs* and *proABC* genes with corresponding locus tags in three widely studied *S. pneumoniae* strains. **Table 5.2** summarises the results from three independent genome-wide screening studies, employing different approaches for *S. pneumoniae* virulence determinants, which found *fhs* and *proABC* to be potential virulence factors.

**Table 5.1:** Description of *S. pneumoniae* genes that were deleted to generate  $\Delta fhs$  and  $\Delta proABC$  *S. pneumoniae* mutants. The corresponding locus tags of the genes in *S. pneumoniae* strains TIGR4 serotype 4) (Tettelin *et al.* 2001), D39V (serotype 2) (Slager, Aprianto, and Veening 2018), and R6 (unencapsulated serotype 2) (Hoskins *et al.* 2001) are shown.

Gene	Product	TIGR4 (Tettelin <i>et al.</i> 2001)	D39V (Slager, Aprianto, and Veening 2018)	R6 (Hoskins <i>et al.</i> 2001)
<i>fhs</i>	formate-- tetrahydrofolate ligase	<i>SP_1229</i>	<i>SPV_1087</i>	<i>spr1109</i>
<i>proB</i>	Glutamate 5-kinase	<i>SP_0931</i>	<i>SPV_0822</i>	<i>spr0832</i>
<i>proA</i>	gamma-glutamyl phosphate reductase	<i>SP_0932</i>	<i>SPV_0823</i>	<i>spr0833</i>
<i>proC</i>	pyrroline-5-carboxylate reductase	<i>SP_0933</i>	<i>SPV_0824</i>	<i>spr0834</i>

**Table 5.2:** Gene fitness results for *fhs* and *proABC* from three mutant screening studies for pneumococcal virulence determinants. van Opijnen and Camilli (van Opijnen and Camilli 2012) used Tn-seq for TIGR4 in 17 *in vitro* and two *in vivo* (mice) conditions and based on contribution to fitness classified genes as follows. (1) Essential: required for growth; (2) Core: fitness defect in  $\geq 13$  conditions; (3) Responsive: phenotype in at least one condition; and (4) Unresponsive: no response in any of the sampled conditions. Liu *et al* (Liu *et al.* 2021) used CRISPRi-Seq to screen D39V strain for *in vitro* and *in vivo* (mice) conditions. Significant reduction in abundance of single guide RNAs (sgRNA) targeting specific genes or operons upon dCas9 induction were categorised as essential genes or operons. sgRNAs that did not change were classified as neutral. Smith *et al* (Smith *et al.* 2021) used Tn-Seq to screen D39W for pneumococcal genes that confer improved fitness during co-infection with IAV *in vivo* in murine models of pneumonia and sepsis. Thirty-two genes including *fhs* and *proABC* were identified by the Smith *et al* study.

Gene	Fitness by Tn-seq in TIGR4 Strain (van Opijnen and Camilli 2012)	Fitness by CRISPRi-Seq in D39V strain (Liu <i>et al.</i> 2021)	Fitness by Tn-seq in D39W Strain (Smith <i>et al.</i> 2021)
<i>fhs</i>	Responsive	Essential for IAV co-infection <i>in vivo</i> , essential for growth in C+Y laboratory medium.	Confers fitness in co-infection with IAV
<i>proB</i>	Responsive - <i>in vivo</i>	Essential for IAV co-infection <i>in vivo</i> .	Confers fitness in co-infection with IAV. Attenuates pneumococcal lethality.
<i>proA</i>	Responsive - <i>in vivo</i>	Essential for IAV co-infection <i>in vivo</i> .	Confers fitness in co-infection with IAV.
<i>proC</i>	Responsive - <i>in vivo</i>	Neutral for IAV co-infection <i>in vivo</i> .	Confers fitness in co-infection with IAV.

IAV = Influenza A virus

### 5.2.2 Mutant construction

The mutant *S. pneumoniae* strains and bacterial RNA-Seq data were generated by Dr Elisa Sevillano-Ramos in Prof. Jeremy Brown's group. The protocol used to generate the *S. pneumoniae* mutants have been described in detail by Sevillano-Ramos *et al* (Ramos-Sevillano, Ercoli, Guerra-Assunção, *et al.* 2021). Briefly, mutants with complete deletion of either *fhs* or *proABC* were constructed in clinical *S. pneumoniae* strain BHN418 (serotype 6B) using overlap extension PCR with custom primers. The constructs were then transformed by homologous recombination and allelic replacement of the target genes with the antibiotic cassette using competence stimulating peptides CSP-1 and CSP-2 (Havarstein, Coomaraswamy, and Morrison 1995).

### 5.2.3 Bacterial growth and transcriptomic experiment design

Purified single colonies of *S. pneumoniae* strains were cultured at 37°C and 5% CO<sub>2</sub> in solid Columbia agar supplemented with 5% horse blood. The cultures were grown to mid-logarithm phase (OD<sub>600</sub> 0.4 - 0.45) in Todd-Hewitt broth supplemented with 0.5% yeast (THY). The cultures were then centrifuged at 12,000g for 8 minutes and bacterial pellets were resuspended in either a fresh preparation of THY or undiluted human serum (obtained from laboratory volunteers) and incubated at 37°C and 5% CO<sub>2</sub> for a further 60 minutes. The cultures were pelleted again as described above and stored at -80°C in RNA protect reagent (Qiagen) following manufacturer's instructions. For each condition (THY or serum), data for at least three independent biological replicates for the mutants ( $\Delta fhs$  and  $\Delta proABC$ ) and WT (strain BHN418) were generated. Total RNA was extracted using the Mirvana miRNA Isolation Kit (Thermo Fisher) following the manufacturer's instructions. DNA was digested and removed using TURBO DNA-free™ Kit (Thermo Fisher) and the product was purified using the RNeasy MinElute Cleanup Kit (Qiagen) and the manufacturer's protocol. Total RNA quality and quantity was assessed by Nanodrop (Thermo Fisher) and the Agilent Bioanalyzer RNA 6000 Nano reagents and Chips.

### 5.2.4 Library preparation and sequencing

Total RNA from each replicate was depleted of ribosomal RNA (rRNA) using the Ribo-Zero Plus rRNA Depletion kit (Illumina). rRNA depleted RNA was used to prepare libraries using



the KAPA RNA HyperPrep kit (Roche) and the in-house protocol from the Pathogen Genomics Unit (PGU), University College London. Libraries were multiplexed to 24 samples per run and sequenced as single end 75bp reads on the NextSeq high output over a 75-cycle run.

### 5.2.5 Quality control, mapping, and quantification of transcripts

Trimmed reads were checked by FastQC v0.11.5, Babraham Bioinformatics, UK (S. 2010) and multiQC v1.9 (Ewels *et al.* 2016) to confirm quality of remaining reads which were then mapped to the reference genomes; (i) the parent serotype 6B genome sequence (BHN418, Accession: NZ\_ASHP01000040.1)(Browall *et al.* 2014) and (ii) KEGG annotated serotype 6B genome sequence (670-6B, Accession: CP002176.1) using bowtie2 v2.4.4 with default settings (Langmead and Salzberg 2012). The generated SAM files were converted to BAM files and sorted by genomic coordinates using SAMtools (Li *et al.* 2009). Mapped reads were visualized in the Integrated Genome Viewer (IGV) (Nicol *et al.* 2009) using the corresponding GTF files of the reference genomes. FeatureCounts v2.0.0 was implemented in a Unix terminal as a bash script to summarize read counts and generate a count matrix of annotated features of the reference genomes in multimapping mode (-M) (Liao, Smyth, and Shi 2014).

### 5.2.6 Differential gene expression analyses

DESeq2 (v1.32.0) normalized counts were transformed using regularized log function within DESeq2 to enable clustering and visualization of expression data. Principal component analysis (PCA) and hierarchical clustering were used to calculate the distance between conditions and replicates of the same condition. Samples that cluster distantly from replicates of the same condition were considered outliers and removed from further analysis. In such cases, the DESeq2 pipeline was rerun without the outlier samples. Differential gene expression was performed and transcripts with expression  $\log_2$  fold change (LFC)  $>1.5$  and an adjusted p-value (false discovery rate [FDR])  $<0.05$  were categorized as differentially expressed. P-values were adjusted for multiple testing in DESeq2 by implementing the Benjamini-Hochberg method. Shrinkage of effect sizes ( $\log_2$

fold change estimates) was implemented with apegglm package within DESeq2 (Zhu, Ibrahim, and Love 2019).

### 5.2.7 Pathway enrichment analysis

Pathway enrichment analysis was performed separately for upregulated and downregulated genes. Pathway over-representation analysis (ORA) of DEGs was performed in R (v3.4.2) with clusterProfiler v4.0.5 (Yu *et al.* 2012). ORA results with adjusted p-value (FDR) <0.05 were categorised as significantly enriched. The reported p-values were adjusted for multiple testing using the Benjamini-Hochberg methods.

### 5.2.8 Identification of correlated gene modules

Differentially expressed genes can be co-ordinately regulated across conditions which can provide powerful biological insights in identifying significant modular patterns (Aprianto *et al.* 2018; Abu-Jamous and Kelly 2018). In order to identify sets of genes that are correlated in gene expression, correlation-based clustering methods and heatmap visualisations were used to reveal co-regulated gene modules (Eisen *et al.* 1998). Two unsupervised classification methods were used to identify co-expressed gene clusters. First, the NbClust R package algorithm (Charrad *et al.* 2014) was applied to a matrix of DEGs to determine the optimal number of clusters by varying all combinations of clusters, distance measures, and clustering methods. The optimal number of clusters proposed by NbClust was then used in a heatmap of hierarchical clustering based on z-scores of transcripts counts across conditions to segregate co-regulated gene modules (Eisen *et al.* 1998). A noted drawback of hierarchical clustering methods is that all genes in a list are assigned into clusters, and this may not reflect underlying biological properties of co-expressed genes (Abu-Jamous and Kelly 2018). Thus, the Clust algorithm was also applied as a comparator for gene module identification (Abu-Jamous and Kelly 2018). Clust is a fully automated command-line program that implements a hybrid binarization of consensus partition matrices (Bi-CoPaM) approach to clustering. Clust does not attempt to assign all genes in a list to a finite set of modules/clusters and has been demonstrated to outperform widely used clustering methods in identifying modules that match biological expectations (Abu-Jamous and Kelly 2018). Modules identified by hierarchical

clustering were named numerically (e.g., Cluster 1, Cluster 2) while modules identified by Clust were name alphabetically (e.g., Cluster A, Cluster B). To verify whether gene modules identified by the two clustering methods were enriched for specific biological pathways, ORA pathway enrichment was performed for gene lists in each identified module.

### 5.2.9 Statistical analyses of genesets

For selected pathways that were found to be significantly enriched in the mutant strains when grown in serum, differences in expression of all the genes within the pathway was assessed and compared to expression levels in growth in THY. DESeq2 Normalised transcript expression counts for all biological replicates in selected pathways were analysed by unpaired t-test to compare between groups of strains and growth conditions. The generated p-values were corrected for multiple testing by the Bonferroni correction method. Adjusted p-values <0.05 were considered to be significant. Ggplot2, rstatix, and ggpubr packages in R (v3.4.2) were used for statistical testing and visualisation of results.

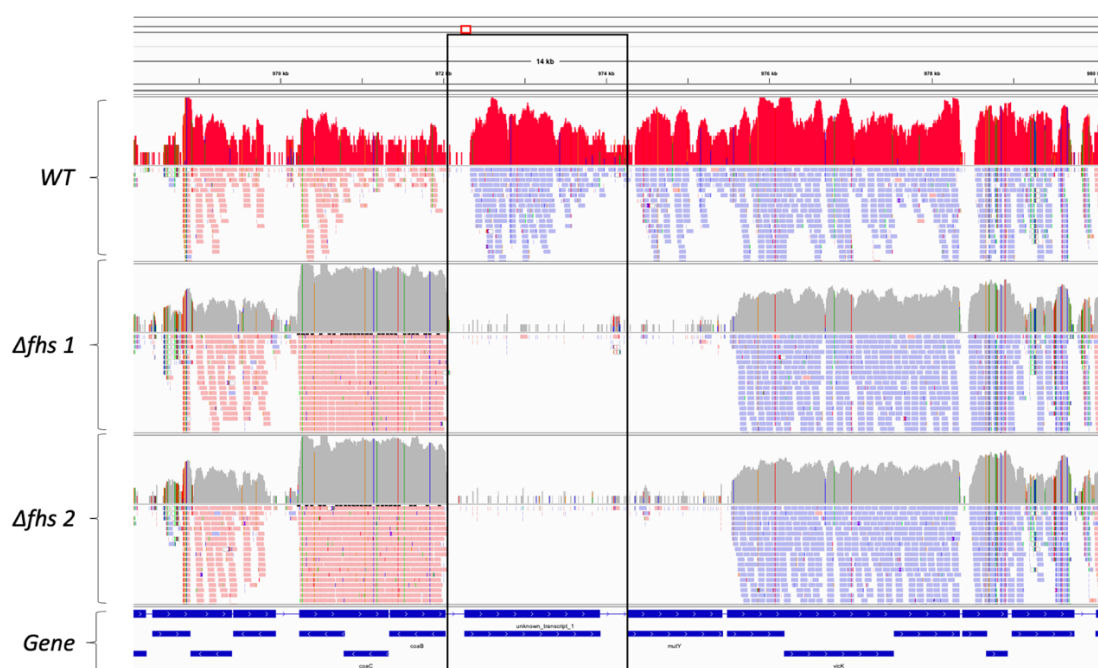
## 5.3 Results

### 5.3.1 *S. pneumoniae* transcriptome sequencing and mapping of reads

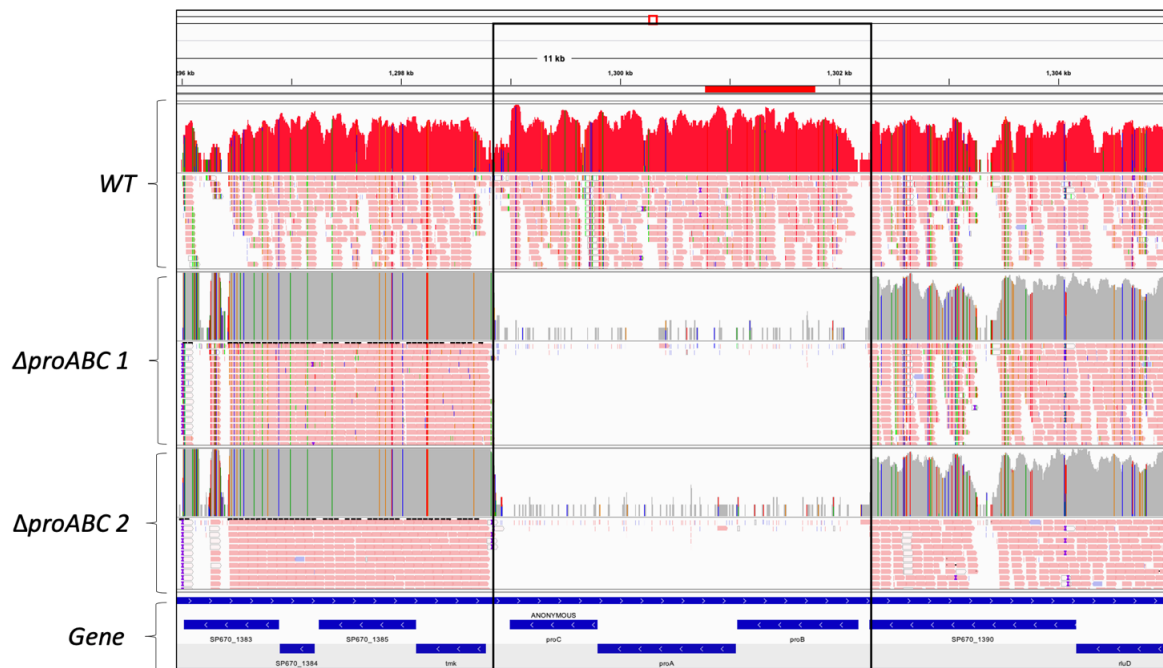
The  $\Delta fhs$  and  $\Delta proABC$  *S. pneumoniae* transcriptome data, generated by Dr Elisa Sevillano Ramos, were aligned to *S. pneumoniae* serotype 6B WT reference genomes BHN418 and SP670-6B. An average of approximately 17,280,850 million 75-bp single end reads (15,555,394 – 19,823,211) were generated for each library (n = 24). Following quality filtering and adapter removal, an average of 98% of reads were retained (97.1% - 98.7%). On average, approximately 15,185,837 million reads per library were successfully mapped to the reference genomes of BHN418 and SP670-6B. Comparison of genome mapping results between BHN418 (Browall *et al.* 2014) and SP670-6B were almost identical (within 0.6% mapping difference for all libraries). To enable further downstream analyses, the mapping results of the KEGG annotated serotype 6B genome SP670-6B (GenBank accession number ASM14709v1) was used for all subsequent analyses.

### 5.3.2 $\Delta fhs$ and $\Delta proABC$ *S. pneumoniae* gene knockout verification

Deletion of the *fhs* and *proABC* genes were verified by aligning and comparing the whole genome sequences of the mutants to the WT. Mapping of transcript reads to the SP670-6B WT genome sequence confirmed that  $\Delta fhs$  and  $\Delta proABC$  *S. pneumoniae* lack transcripts mapping to the coding regions of *fhs* and *proABC* respectively, in contrast to reads mapped to the WT (**Figure 5.1** and **Figure 5.2**). The deletion of *fhs* however, also seems to have affected transcription of the downstream gene *mutY* (**Figure 5.1**).



**Figure 5.1:** IGV visualisation of transcripts mapped to *S. pneumoniae* serotype 6B WT genome sequence (SP-670) confirmed deletion of the *fhs* gene in  $\Delta fhs$  *S. pneumoniae* as indicated by the lack of transcripts mapping to the *fhs* exon. In contrast, *fhs* transcripts mapped to the *fhs* coding region in WT. Log densities (top panels in each strain) of reads from one WT replicate in serum (red shaded) and two mutant replicates (grey shaded) are shown. The bottom panels (in each strain) show the absolute read counts with purple colour showing reads mapped to the forward strand and pink colour showing reads mapped to the reverse strand. The black rectangle highlights *fhs* exon (coding sequence) in WT and mutant. The bottom (Gene) row shows annotation of the highlighted genomic region with information on strand orientation and gene names. The transcription of the downstream gene *mutY* appear to be also affected by deletion of *fhs*.

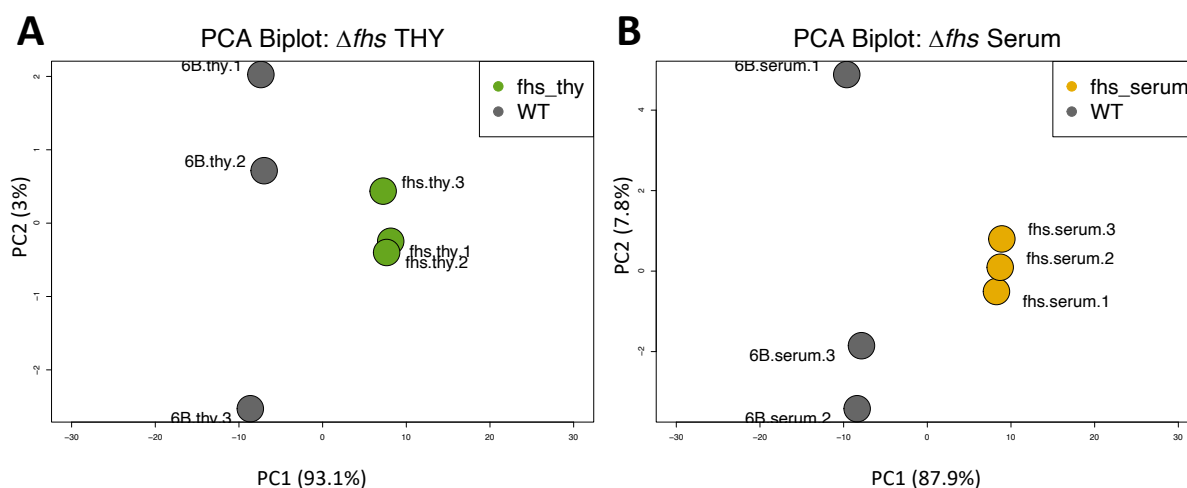


**Figure 5.2:** IGV visualisation of transcripts mapped to *S. pneumoniae* serotype 6B WT genome sequence (SP-670) confirmed deletion of the *proABC* genes in  $\Delta proABC$  *S. pneumoniae* as indicated by lack of transcripts mapping to the *proABC* exons. In contrast, *proABC* transcripts mapped to the *proABC* coding regions in WT. Log densities (top panels in each strain) of reads from one WT replicate in serum (red shaded) and two mutant replicates in serum (grey shaded) are shown. The bottom panels (in each strain) show the absolute read counts with pink colour indicating reads mapped to the reverse strand. The black rectangle highlights *proABC* exons (coding sequences) in WT and mutant. The bottom (Gene) row shows annotation of the highlighted genomic region with information on strand orientation and gene names.

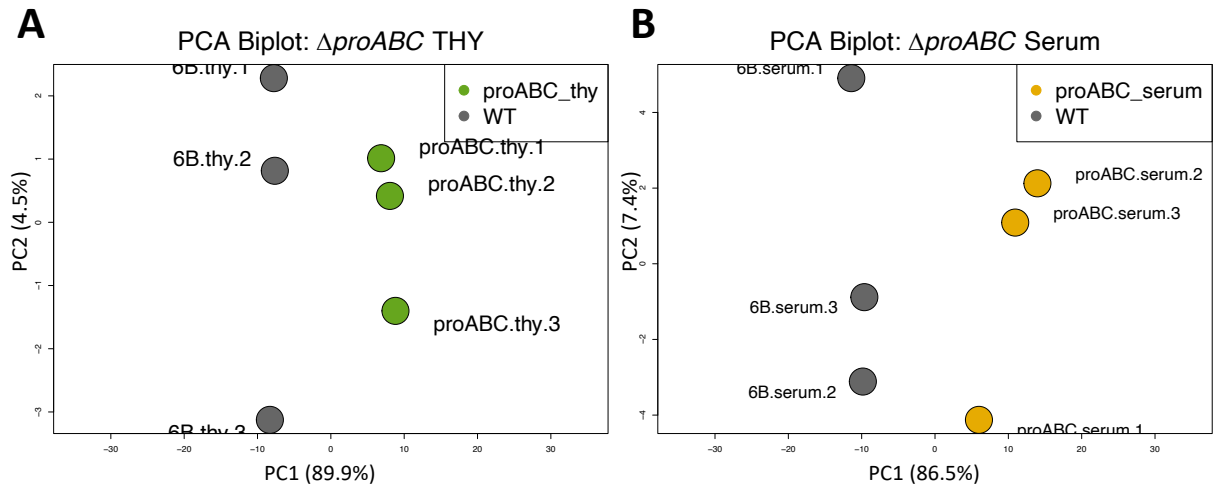
### 5.3.3 $\Delta fh$ s and $\Delta proABC$ *S. pneumoniae* transcriptome data quality

Following mapping, normalisation, and transformation of raw transcript reads for all libraries, principal components analyses were performed to assess the quality of data generated. The first two components showed high reproducibility between the biological replicates of the two media conditions and showed clear separation of the mutants from WT (**Figure 5.3** and **Figure 5.4**). The first two principal components accounted for >90% of the cellular variability between mutants and the WT (**Figure 5.3** and **Figure 5.4**). We

proceeded to identify transcripts that were differentially expressed in each mutant relative to the WT in THY and serum.



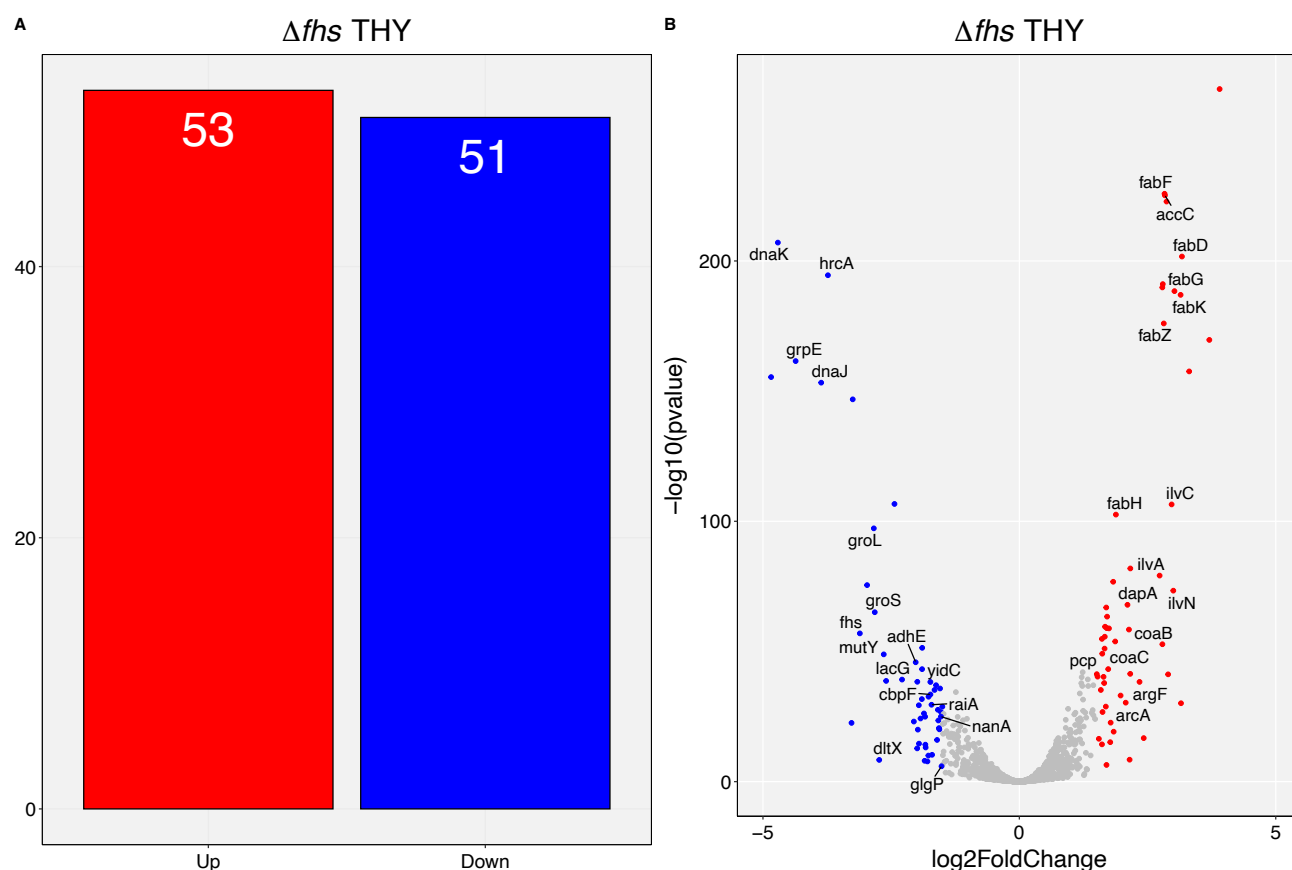
**Figure 5.3: Principal component analysis (PCA) of  $\Delta fhs$  *S. pneumoniae* and WT transcriptomes.** RNA-Seq data was generated from  $\Delta fhs$  and WT *S. pneumoniae*. Bacterial strains were grown to mid-log phase and transferred to either Todd-Hewitt broth + Yeast (THY) medium or human serum. **(A)** PCA showing clustering of  $\Delta fhs$  (green; n= 3 biological replicates) and WT (grey; n=3 replicates) *S. pneumoniae* grown in THY. **(B)** PCA showing clustering of  $\Delta fhs$  (gold; n= 3 biological replicates) and WT (grey; n=3 replicates) *S. pneumoniae* grown in human serum.



**Figure 5.4: Principal component analysis (PCA) of  $\Delta proABC$  *S. pneumoniae* and WT transcriptomes.** RNA-Seq data was generated from  $\Delta proABC$  and WT *S. pneumoniae*. Bacterial strains were grown to mid-log phase and transferred to either Todd-Hewitt broth + Yeast (THY) medium or human serum. **(A)** PCA showing clustering of  $\Delta proABC$  (green; n= 3 biological replicates) and WT (grey; n=3 replicates) *S. pneumoniae* grown in THY. **(B)** PCA showing clustering of  $\Delta proABC$  (gold; n= 3 biological replicates) and WT (grey; n=3 replicates) *S. pneumoniae* grown in human serum.

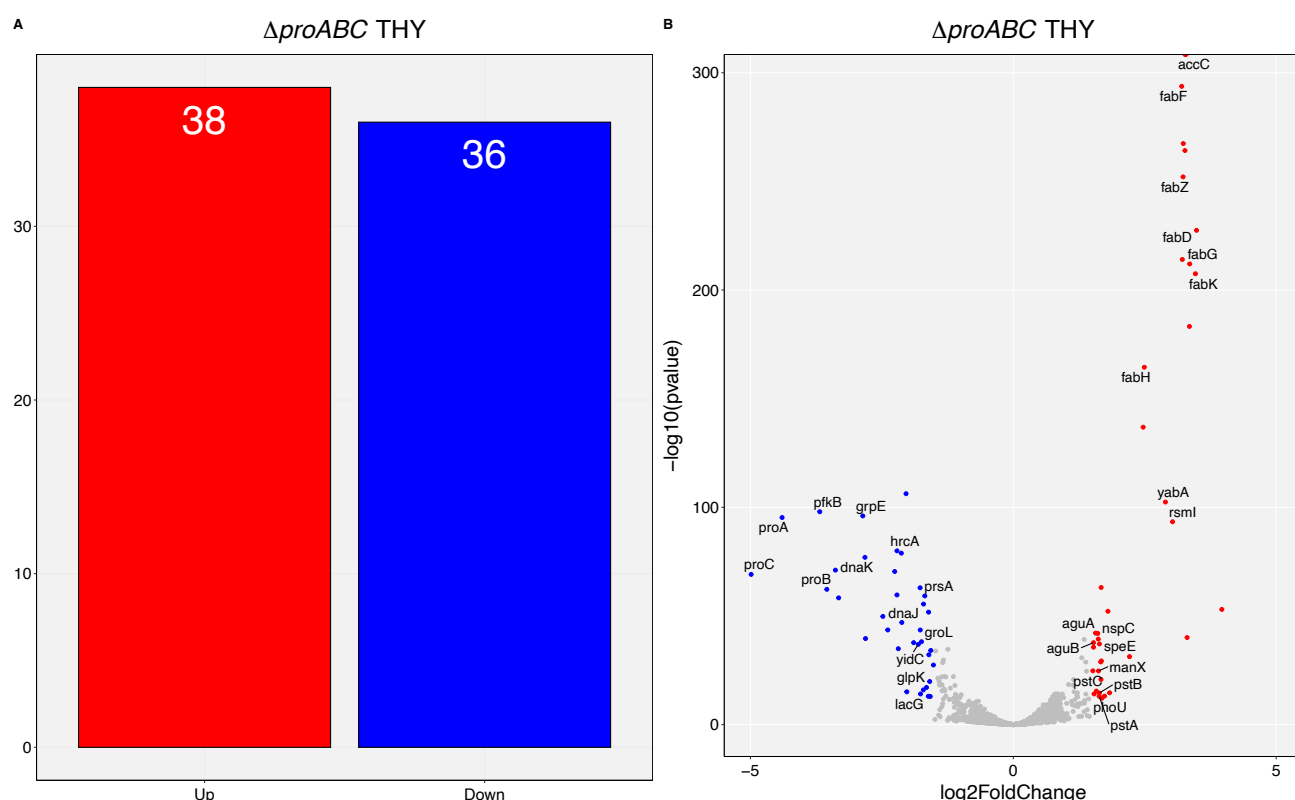
#### 5.3.4 Modest differences in the transcriptome of $\Delta fhs$ and $\Delta proABC$ *S. pneumoniae* compared to the WT in THY

THY is a serum-free liquid culture that is nutritionally enriched with high glucose, amino acid, and protein content (Todd and Hewitt 1932; Sanchez-Rosario and Johnson 2021). This nutritional environment makes THY an ideal growth medium for fastidious beta-haemolytic *Streptococci* such as *S. pneumoniae* (Todd and Hewitt 1932; Sanchez-Rosario and Johnson 2021). The total number of up- and down-regulated genes of each of  $\Delta fhs$  (**Figure 5.5A**) and  $\Delta proABC$  (**Figure 5.6A**) *S. pneumoniae* were very similar when grown in THY medium compared to the WT.



**Figure 5.5: Differential gene expression of  $\Delta fhs$  *S. pneumoniae* grown in Todd-Hewitt broth + Yeast (THY) medium. (A)** Bar plots show the total number of differentially expressed genes (DEGs) that significantly increased (red) or reduced (blue) expression in  $\Delta fhs$  *S. pneumoniae* relative to WT. **(B)** Volcano plot of  $\Delta fhs$  *S. pneumoniae* DEGS relative to WT in THY. Red dots represent upregulated genes and blue coloured dots represent downregulated genes. Grey coloured dots are transcripts that did not meet DEGs threshold. Short gene names of DEGs are annotated on the plot when available. DEGs were defined as genes with  $\log_2$  fold change  $>1.5$  and  $\text{FDR} < 0.05$ .





**Figure 5.6: Differential gene expression of  $\Delta proABC$  *S. pneumoniae* grown in Todd-Hewitt broth + Yeast (THY) medium. (A)** Bar plots show the total number of differentially expressed genes (DEGs) that significantly increased (red) or reduced (blue) expression in  $\Delta proABC$  *S. pneumoniae* relative to WT. **(B)** Volcano plot of  $\Delta proABC$  *S. pneumoniae* DEGS relative to WT in THY. Red dots represent upregulated genes and blue coloured dots represent downregulated genes. Grey coloured dots are transcripts that did not meet DEGs threshold. Short gene names of DEGs are annotated on the plot when available. DEGs were defined as genes with  $\log_2$  fold change  $>1.5$  and FDR  $<0.05$ .

### 5.3.5 $\Delta fhs$ and $\Delta proABC$ *S. pneumoniae* upregulate expression of fatty-acid synthesis genes in THY

A common theme in the gene expression of  $\Delta fhs$  and  $\Delta proAB$  *S. pneumoniae* in THY was the increased expression of fatty acid synthesis (FAS) genes (**Figure 5.5B** and **Figure 5.6B**) that are located within an operon and co-ordinately regulated in *S. pneumoniae* (Jerga and Rock 2009). In  $\Delta fhs$  *S. pneumoniae*, all but one (SP670\_0492, LFC = 1.62, p-adj = 0.12) of the thirteen FAS genes were significantly upregulated in THY relative to WT. All thirteen genes in the FAS operon (SP670\_0489, SP670\_0490, *fabH*, SP670\_0492, *fabK*, *fabD*, *fabG*,

*fabF*, *SP670\_0497*, *fabZ*, *accC*, *SP670\_0500*, *SP670\_0501*) were differentially upregulated by  $\Delta$ *proAB* *S. pneumoniae*. The functions of *fhs* and *proAB* genes are incompletely understood but have not been shown to be involved in biological processes related to fatty-acid biosynthesis (Ramos-Sevillano, Ercoli, Felgner, *et al.* 2021; van Opijnen and Camilli 2012; Crowley *et al.* 1997; Mahdi *et al.* 2012; Afzal, Shafeeq, and Kuipers 2016). Therefore, upregulation of FAS genes in  $\Delta$ *fhs* and  $\Delta$ *proABC* may be an indirect effect of mutation. In both  $\Delta$ *proAB* and  $\Delta$ *fhs* *S. pneumoniae*, the FAS genes were statistically the most significantly upregulated transcripts (lowest adjusted p-values) in THY media (**Figure 5.5B** and **Figure 5.6B**), implying that the quantitative differences between the biological groups was relatively consistent within each group of replicates.

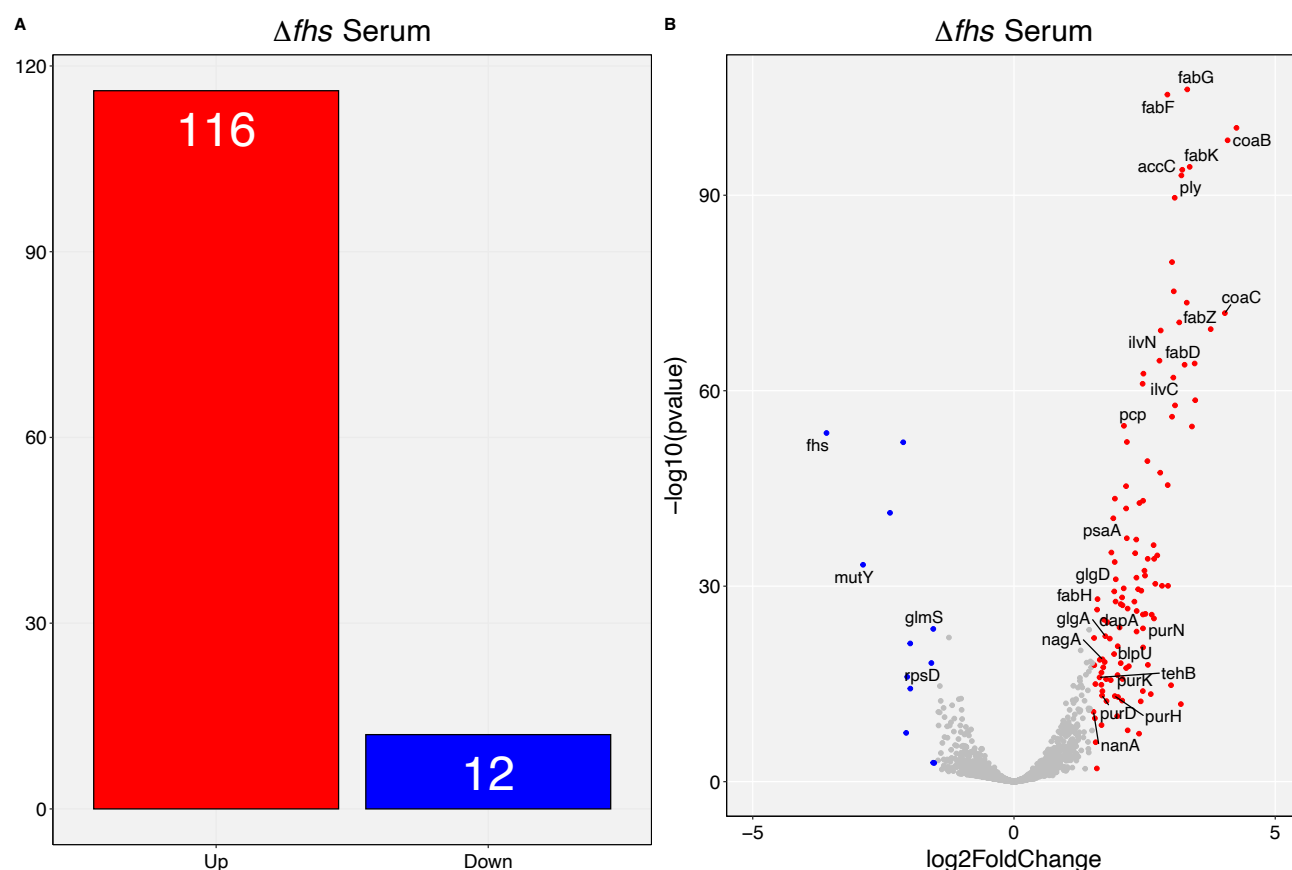
### 5.3.6 $\Delta$ *fhs* and $\Delta$ *proABC* *S. pneumoniae* upregulate expression of metabolic genes in THY

The upregulated transcripts by  $\Delta$ *fhs* and  $\Delta$ *proAB* *S. pneumoniae* in THY were dominantly involved in bacterial metabolism and biosynthesis (**Figure 5.5B** and **Figure 5.6B**). For  $\Delta$ *fhs* *S. pneumoniae*, genes involved in amino acid synthesis and acquisition such as *ilvA*, *ilvC*, *ilvN*, *SP670\_0519* were distinctly upregulated as well as the essential pantothenate and coenzyme A (coA) biosynthesis pathway genes *coaB* and *coaC* (**Figure 5.5B**). coA is an essential co-factor and regulator of key metabolic enzymes in several cellular pathways with up to 9% of all bacterial enzymes estimated to use this co-factor (Spry, Kirk, and Saliba 2008). The *coaB* and *coaC* genes are immediately upstream of the *fhs* gene and their upregulation suggests that *coaB* and *coaC* may be directly involved in *fhs* biosynthesis.  $\Delta$ *proABC* *S. pneumoniae* upregulated expression of pneumococcal genes involved in carbohydrate utilisation and biosynthesis (*manX*, *SP670\_2250*, *SP670\_0354*, *SP670\_0355*, *SP670\_2248*, *SP670\_2249*, *SP670\_2250*), ABC phosphotransferases (*pstA*, *pstB*, *phoU*, *SP670\_1404*) (**Figure 5.6**), and a set of polyamine biosynthesis genes (*aguB*, *aguA*, *nspC*, *SP670\_1400*, *speE*, *SP670\_1404*) (**Figure 5.6**) that constitute a single operon (Ayoola *et al.* 2019).

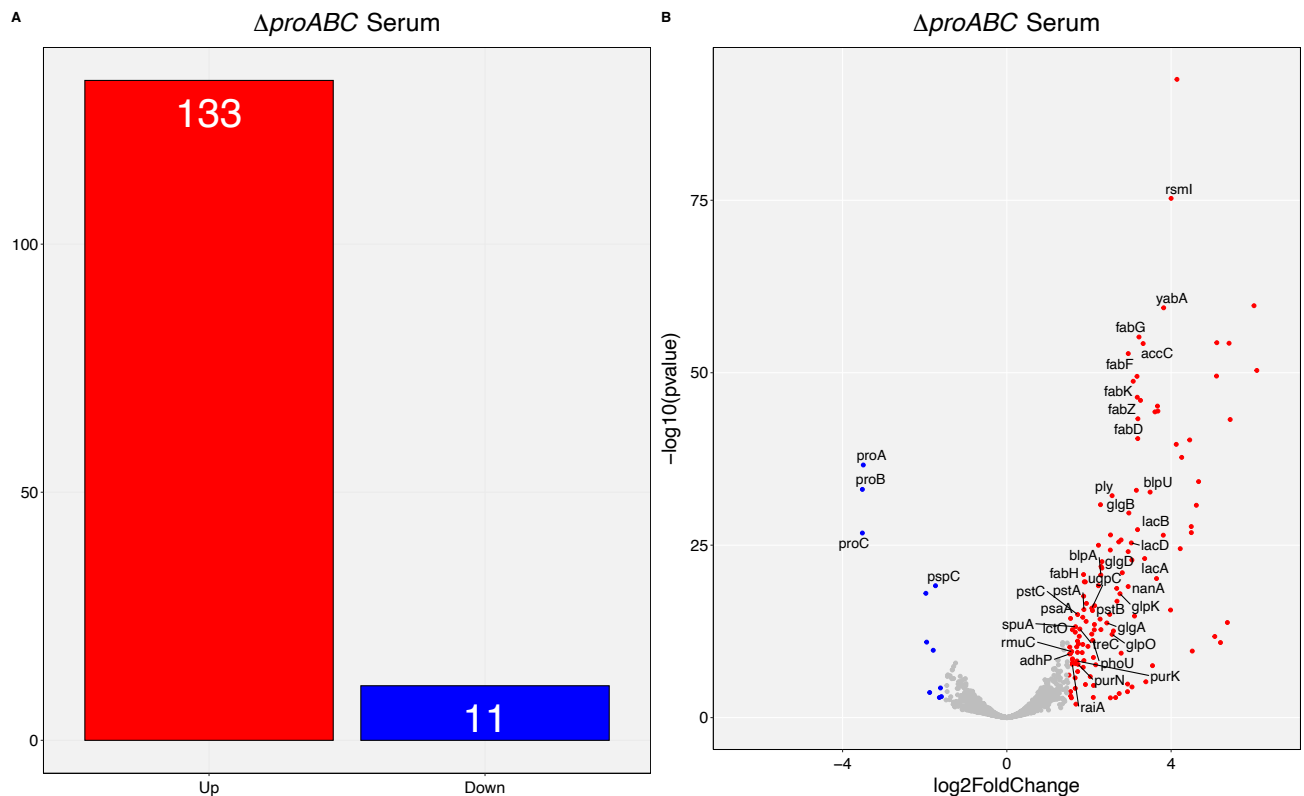
The commonly downregulated genes between  $\Delta$ *fhs* and  $\Delta$ *proABC* *S. pneumoniae* (i.e., genes that were more highly expressed by the WT) in THY included stress response genes such as chaperon proteins *groEL*, *dnaJK*, and the chaperon regulator *hrcA* (**Figure 5.5B** and **Figure 5.6B**).

### 5.3.7 Growth of $\Delta fhs$ and $\Delta proABC$ *S. pneumoniae* in human serum to induce stress results in a distinct virulence gene expression pattern compared to growth in THY

Serum is a nutrient limited growth medium and may more closely mimic the stress environment encountered by *S. pneumoniae in vivo* compared to THY. Unlike growth in THY, the growth of  $\Delta fhs$  and  $\Delta proABC$  *S. pneumoniae* in serum was poor compared to the WT. In contrast to the modest differences in the number of up- and down-regulated genes when grown in THY,  $\Delta fhs$  and  $\Delta proABC$  *S. pneumoniae* growth in serum resulted in a large number of upregulated genes. There were 10 times as many upregulated genes as downregulated genes for  $\Delta fhs$  *S. pneumoniae* (**Figure 5.7A**) and 12 times as many upregulated genes as downregulated genes for  $\Delta proABC$  *S. pneumoniae* in serum (**Figure 5.8A**). In addition, several well-known pneumococcal virulence genes were upregulated in both mutants in serum suggesting that serum culture necessitates an adaptive stress response by the mutants (**Figures 5.7B** and **Figure 5.8B**). Subsequent functional analyses were focused on delineating gene expression results of the mutants in the serum culture conditions.



**Figure 5.7: Differential gene expression of  $\Delta fhs$  *S. pneumoniae* grown in human serum.** **(A)** Bar plots show the total number of differentially expressed genes (DEGs) that have significantly increased (red) or reduced (blue) expression in  $\Delta fhs$  *S. pneumoniae* relative to WT in human serum. **(B)** Volcano plot of  $\Delta fhs$  *S. pneumoniae* DEGs relative to WT in human serum. Red dots represent upregulated genes and blue coloured dots represent downregulated genes. Grey coloured dots are transcripts that did not meet DEGs threshold. Short gene names of DEGs are annotated on the plot when available. DEGs were defined as genes with  $\log_2$  fold change  $>1.5$  and FDR  $<0.05$ .



**Figure 5.8: Differential gene expression of  $\Delta proABC$  *S. pneumoniae* grown in human serum (A)** Bar plots show the total number of differentially expressed genes (DEGs) that have significantly increased (red) or reduced (blue) expression in  $\Delta proABC$  *S. pneumoniae* relative to WT in human serum. **(B)** Volcano plot of  $\Delta fhs$  *S. pneumoniae* DEGs relative to WT in human serum. Red dots represent upregulated genes and blue coloured dots represent downregulated genes. Grey coloured dots are transcripts that did not meet DEGs threshold. Short gene names of DEGs are annotated on the plot when available. DEGs were defined as genes with  $\log_2$  fold change  $>1.5$  and FDR  $<0.05$ .

### 5.3.8 Deletion of *fhs* or *proABC* genes linked to dysregulation of fatty-acid biosynthesis in *S. pneumoniae* in serum

Similar to growth in THY,  $\Delta fhs$  and  $\Delta proABC$  *S. pneumoniae* significantly increased expression of FAS genes in serum, confirming a link between the mutations and regulation of FAS genes in this dataset (**Figure 5.7B** and **Figure 5.8B**).  $\Delta fhs$  and  $\Delta proABC$  *S. pneumoniae* significantly upregulated the expression of all thirteen FAS pathway genes in serum. Lactobacillales such as *S. pneumoniae* respond to exogenous fatty acids by suppressing *de novo* biosynthesis pathway and exclusively utilizing the available

extracellular fatty acids for membrane phospholipid synthesis (Parsons *et al.* 2011; Brinster *et al.* 2009; Yao and Rock 2017). Human serum is a rich source of fatty acids ( $\sim 3\text{g l}^{-1}$ ) which we expect would subdue endogenous biosynthesis of fatty acids and hence downregulation of the FAS genes (Brinster *et al.* 2009). Thus, the upregulation of FAS genes may be an intrinsic property of  $\Delta fhs$  and  $\Delta proABC$  *S. pneumoniae*. We reasoned that the deletion of the *fhs* and *proABC* genes may influence *S. pneumoniae* sensing and regulation of exogenous fatty acids metabolism and potentially other unrelated pathways.

### 5.3.9 Virulence genes expression by $\Delta fhs$ *S. pneumoniae* in human serum

The top 30 most significantly upregulated genes for  $\Delta fhs$  *S. pneumoniae* by fold change ranking are described in **Table 5.3** (excluding FAS genes). Importantly, culture of  $\Delta fhs$  *S. pneumoniae* in serum induced upregulation of pneumococcal virulence factors, including pneumolysin (*ply*), ABC transporter substrate-binding lipoprotein (*psaA*), and neuraminidase A (*nanA*) (**Figure 5.7B**). Virulence genes *ply*, *psaA*, and *nanA* are all involved in *S. pneumoniae* adherence and colonisation (Weiser, Ferreira, and Paton 2018). *psaA* plays a critical role in regulation of gene expression of *S. pneumoniae* in response to oxidative stress response and intracellular redox homeostasis (Tseng *et al.* 2002) and is also implicated as a promoter of bacterial adhesion to host cells (Rajam *et al.* 2008). Neuraminidase A (*nanA*) enhances adherence of *S. pneumoniae* to the mucosal surfaces and is also involved in biofilm formation (Parker *et al.* 2009; Manco *et al.* 2006). Pneumolysin is a bacterial toxin that promotes biofilm formation (Shak *et al.* 2013), plays a role in efficient colonisation and transmission of *S. pneumoniae* in the nasopharynx (Zafar *et al.* 2017; Hotomi *et al.* 2016), and is involved in regulating the complement system. The increased expression of these virulence genes does not appear to increase virulence of  $\Delta fhs$  *S. pneumoniae* which was attenuated for virulence in murine studies (Ramos-Sevillano, Ercoli, Felgner, *et al.* 2021).

In addition to the upregulation of virulence genes, and consistent with complex adaptation of bacterial pathogens to stress environments, the expressions of several metabolic and biosynthesis genes were upregulated by  $\Delta fhs$  *pneumoniae* in serum.

**Table 5.3:** Top 30 upregulated genes by  $\Delta fhs$  *S. pneumoniae* relative to WT during culture in human serum. Genes were sorted by  $\log_2$  fold change differences. Fatty-acid synthesis (FAS) genes were excluded, and the Rank column indicates ranking of upregulated genes if FAS genes were included.

Rank	$\log_2$ FC	P-adj	Locus ID	Gene	Product
1	4.26	4.53E-101	SP670_RS05185	<i>coaB</i>	phosphopantothenate-cysteine ligase
2	4.09	3.69E-99	SP670_RS05175		ECF transporter S component
3	4.04	1.25E-72	SP670_RS05180	<i>coaC</i>	phosphopantothenoylcysteine decarboxylase
4	3.76	3.57E-70	SP670_RS07645		biotin transporter BioY
5	3.47	2.88E-59	SP670_RS09565		xanthine phosphoribosyltransferase
6	3.46	6.64E-65	SP670_RS10080		hypothetical protein
7	3.41	3.12E-55	SP670_RS13570		ATP-binding cassette domain-containing protein
10	3.31	3.06E-74	SP670_RS09570		purine permease
13	3.20	9.05E-94	SP670_RS10075	<i>ply</i>	cholesterol-dependent cytolysin pneumolysin
14	3.19	1.26E-12	SP670_RS12200		hypothetical protein
19	3.05	9.68E-63	SP670_RS02610	<i>ilvC</i>	ketol-acid reductoisomerase
20	3.03	9.79E-57	SP670_RS10085		hypothetical protein
21	3.03	1.82E-80	SP670_RS07980		peptide ABC transporter substrate-binding protein
22	3.01	1.59E-15	SP670_RS13405		GNAT family N-acetyltransferase
23	2.95	8.80E-31	SP670_RS02620		hypothetical protein
24	2.94	3.07E-46	SP670_RS02615		hypothetical protein
26	2.83	8.71E-31	SP670_RS00620		phosphoribosylaminoimidazolesuccinocarboxamide synthase
27	2.81	5.78E-70	SP670_RS02605	<i>ilvN</i>	acetolactate synthase small subunit
28	2.80	3.83E-48	SP670_RS00625		phosphoribosylformylglycinamidine synthase
29	2.78	2.47E-65	SP670_RS10090		DUF4231 domain-containing protein
30	2.74	1.90E-35	SP670_RS00630		amidophosphoribosyltransferase
31	2.70	4.31E-31	SP670_RS04975		phosphotransferase family protein
32	2.68	6.36E-35	SP670_RS08760		DUF386 domain-containing protein

(Table continued)

33	2.68	9.1E-26	SP670_RS00640	<i>purN</i>	phosphoribosylglycinamide formyltransferase
34	2.67	5.03E-37	SP670_RS04985		NTP transferase domain-containing protein
35	2.64	2.35E-26	SP670_RS03045		CPBP family intramembrane metalloprotease
36	2.62	3.68E-14	SP670_RS03050		hypothetical protein
37	2.56	1.19E-18	SP670_RS12500		hypothetical protein
38	2.56	6.33E-35	SP670_RS08745		YesL family protein
39	2.55	6.38E-50	SP670_RS09180		aromatic acid exporter family protein

### 5.3.10 Virulence gene expression by $\Delta$ *proABC* *S. pneumoniae* in human serum

When sorted by fold changes relative to the wild type, the top genes upregulated by  $\Delta$ *proABC* *S. pneumoniae* were dominated by genes involved in carbohydrate metabolism and utilisation (**Table 5.4**). Similar to  $\Delta$ *fhs* *S. pneumoniae*,  $\Delta$ *proABC* *S. pneumoniae* upregulated expression of pneumococcal virulence genes *ply*, *nanA*, and *psaA* (**Figure 5.8B**) in serum despite also exhibiting an attenuated phenotype in murine models of virulence. To understand the themes of gene expression during serum stress, we conducted biological pathway and module analyses based on DEGs.



**Table 5.4:** Top 30 upregulated genes by  $\Delta proABC$  *S. pneumoniae* relative to the WT when cultured in human serum. Genes were sorted by  $\log_2$  fold change differences.

Rank	$\log_2$ FC	P-adj	Locus ID	Gene	Product
1	6.09	4.63E-51	SP670_RS00695		PTS system mannose/fructose/N-acetylgalactosamine-transporter subunit IIB
2	6.02	1.93E-60	SP670_RS00690		beta-galactosidase
3	5.44	6.09E-44	SP670_RS00700		PTS mannose/fructose/sorbose/N-acetylgalactosamine transporter subunit IIC
4	5.41	5.29E-55	SP670_RS03510		PTS sugar transporter subunit IIB
5	5.37	1.64E-14	SP670_RS00705		PTS mannose/fructose/sorbose transporter family subunit IID
6	5.20	1.32E-11	SP670_RS03505		PTS sugar transporter subunit IIA
7	5.11	4.47E-55	SP670_RS05325		PTS sugar transporter subunit IIA
8	5.11	3.07E-50	SP670_RS03515		PTS galactitol transporter subunit IIC
9	5.06	1.73E-12	SP670_RS00710		PTS sugar transporter subunit IIA
10	4.67	6.09E-35	SP670_RS05330		PTS sugar transporter subunit IIB
11	4.61	1.64E-31	SP670_RS03520		hypothetical protein
12	4.51	2.26E-10	SP670_RS00715		SIS domain-containing protein
13	4.49	1.52E-27	SP670_RS03525		DUF4982 domain-containing protein
14	4.49	2.01E-28	SP670_RS03050		hypothetical protein
15	4.45	5.58E-41	SP670_RS08780		PTS transporter subunit EIIC
16	4.26	1.87E-38	SP670_RS03045		CPBP family intramembrane metalloprotease
17	4.22	3.23E-25	SP670_RS13405		GNAT family N-acetyltransferase
18	4.14	2.93E-93	SP670_RS06865		DNA polymerase III subunit delta'
19	4.12	2.42E-40	SP670_RS08785		N-acetylmannosamine-6-phosphate 2-epimerase
20	4.00	5.26E-76	SP670_RS06855	<i>rsmI</i>	16S rRNA (cytidine(1402)-2'-O)-methyltransferase

(Table continued)

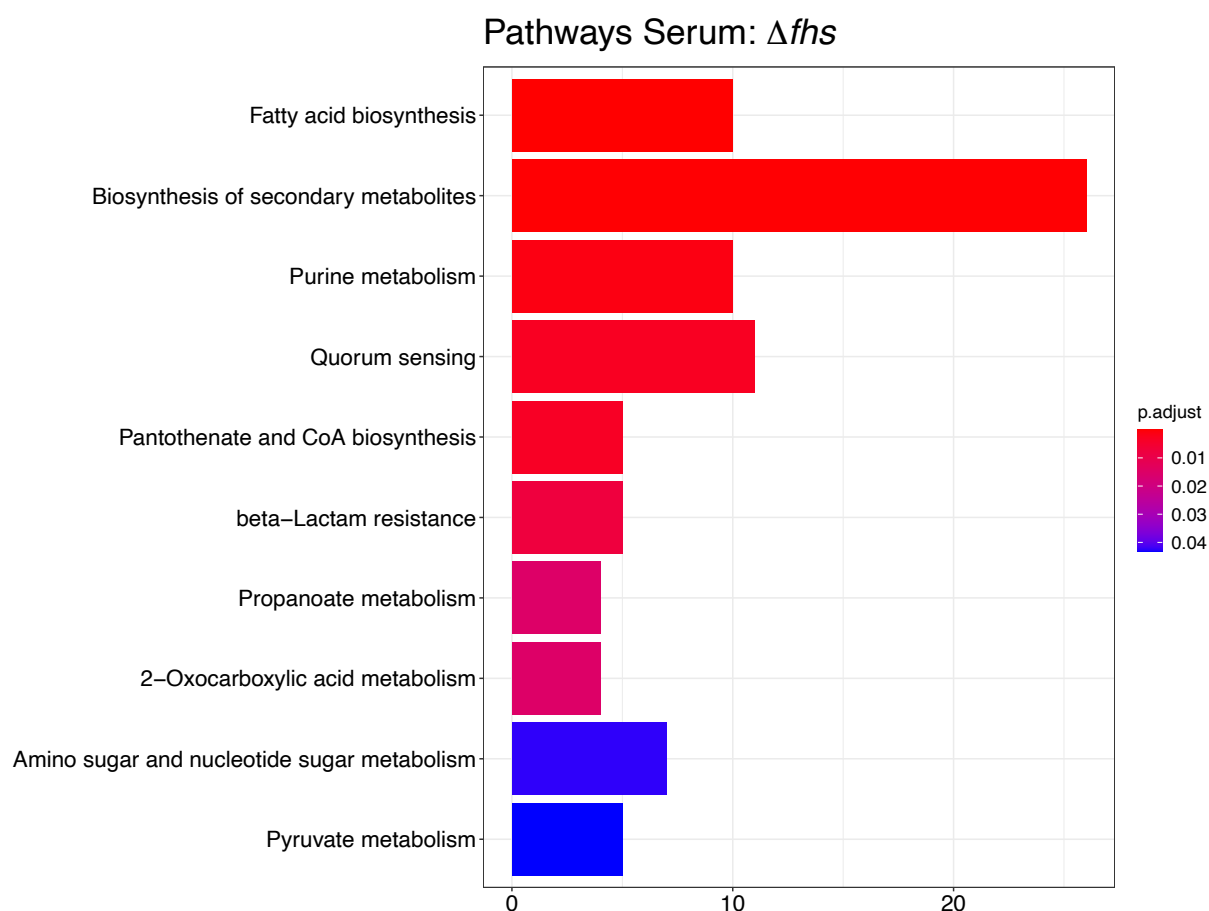
21	3.99	2.53E-16	SP670_RS13400		DNA alkylation repair protein
22	3.82	3.88E-60	SP670_RS06860	<i>yabA</i>	DNA replication initiation control protein YabA
23	3.81	3.57E-27	SP670_RS03055		CPBP family intramembrane metalloprotease
24	3.68	3.66E-45	SP670_RS06870		dTMP kinase
25	3.67	6.89E-46	SP670_RS06750		ABC transporter ATP-binding protein
26	3.64	6.91E-21	SP670_RS08775		carbohydrate ABC transporter substrate-binding protein
27	3.61	4.75E-45	SP670_RS06745		hypothetical protein
28	3.55	2.95E-08	SP670_RS08770		sugar ABC transporter permease
29	3.49	2.02E-33	SP670_RS00595	<i>blpU</i>	bacteriocin-like peptide BlpU
30	3.39	6.40E-06	SP670_RS08760		DUF386 domain-containing protein

### 5.3.11 KEGG pathways enrichment analysis for $\Delta fhs$ and $\Delta proABC$ *S. pneumoniae* in human serum

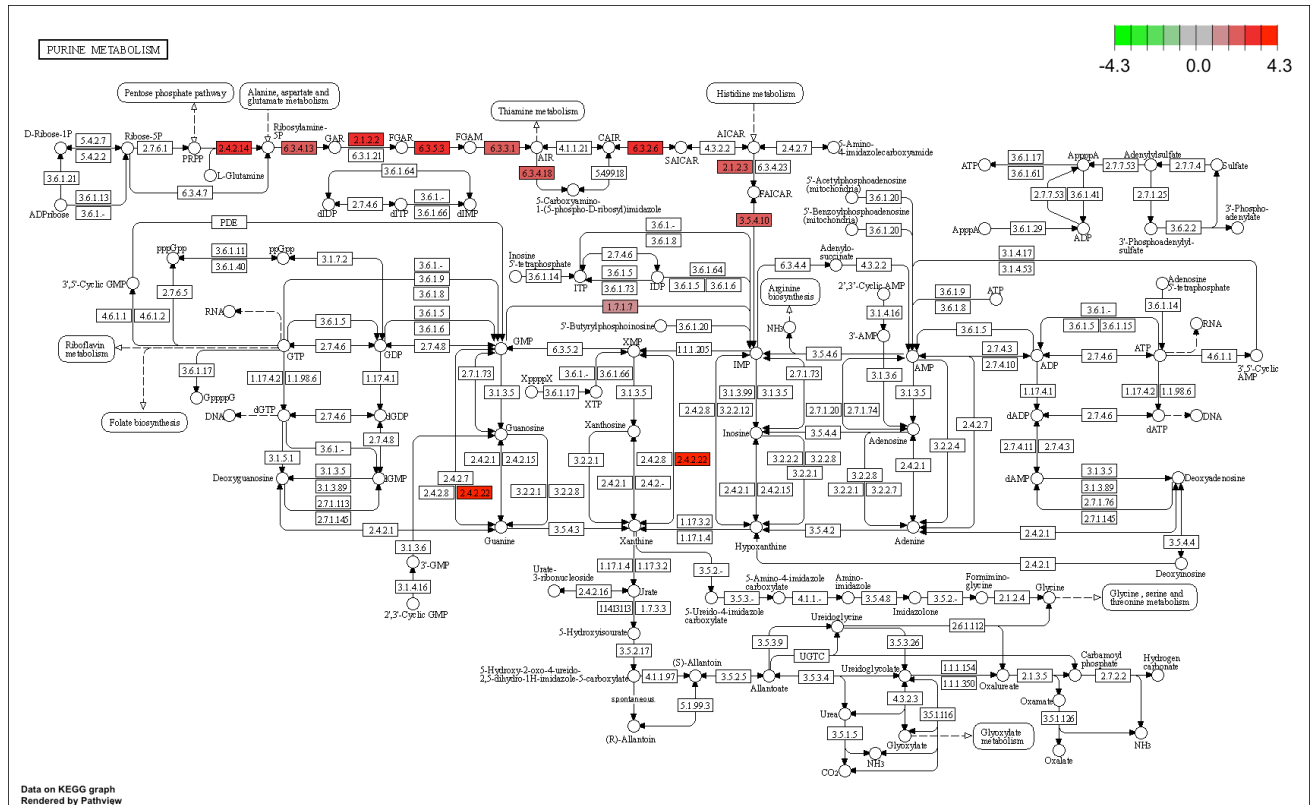
To gain insights into functional mechanisms, we utilised the KEGG database of annotated biological pathways to identify known pathways that were enriched by the upregulated genes of the mutants during culture in serum (Kanehisa 2000). Gene enrichment analyses was performed by over representation analysis (ORA) using KEGG database annotations for *S. pneumoniae* strain SP670-6B.

#### 5.3.11.1 Biological pathways enriched by $\Delta fhs$ *S. pneumoniae* in human serum

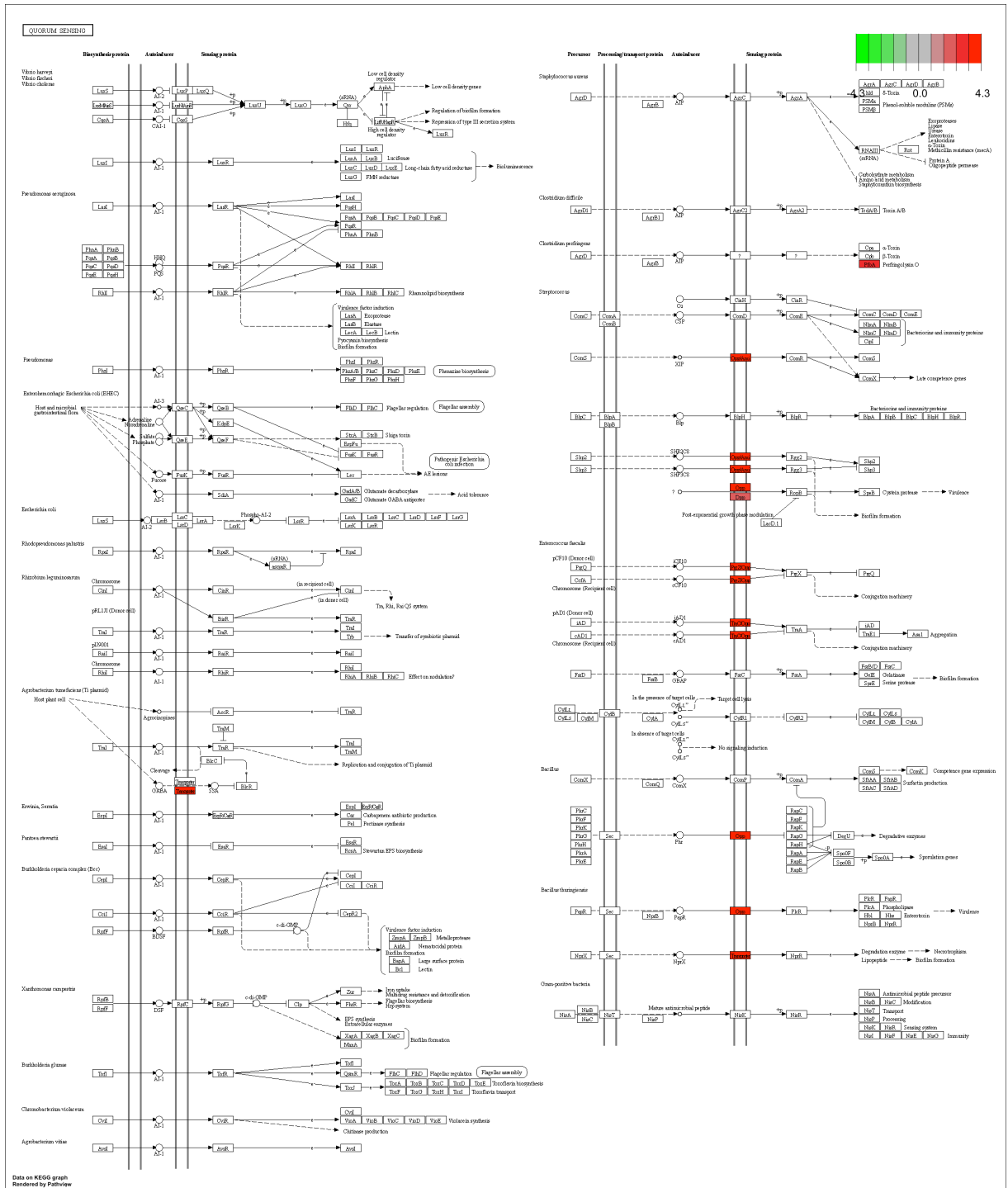
Pathway analyses of  $\Delta fhs$  *S. pneumoniae* suggested significant rewiring of the bacterial biosynthesis pathways in human serum (**Figure 5.9**). Pathways such as purine metabolism (**Figure 5.10**), quorum sensing (**Figure 5.11**), and biosynthesis of secondary metabolites (**Figure 5.12**) were significantly enriched (upregulated) when  $\Delta fhs$  *S. pneumoniae* strain was cultured in serum.



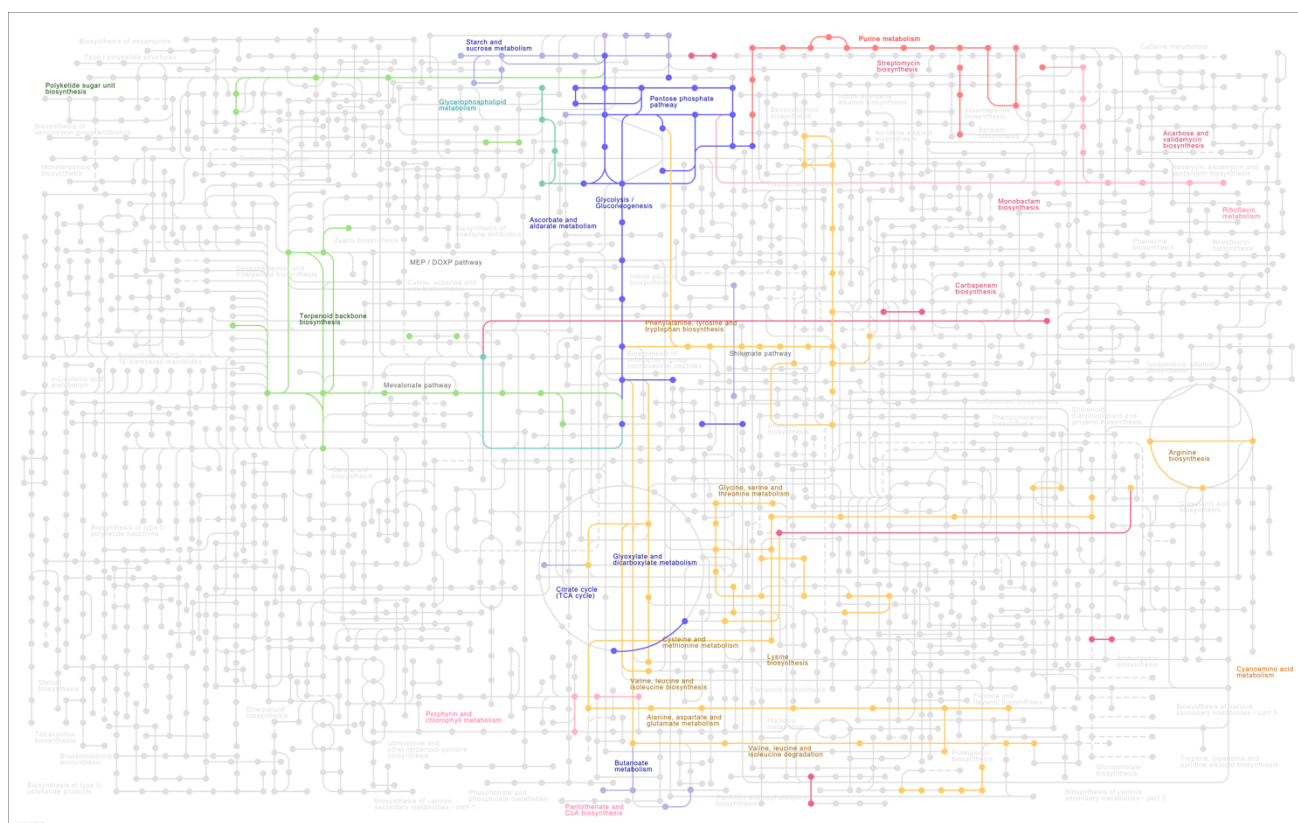
**Figure 5.9: Pathway enrichment analysis of the upregulated genes of  $\Delta fhs$  *S. pneumoniae* relative to WT during culture in human serum.** Pathway enrichment was performed by over representation analysis. Pathways with adjusted p-value  $<0.05$  were considered to be significantly enriched. All pathways shown in the plot were enriched with the length of the bar corresponding to the number of DEGs in the respective pathways. The colour of the bars denotes level of statistical significance based on adjusted p-value with red coloured being the most significantly enriched pathways.



**Figure 5.10: Visualisation of the Purine metabolism KEGG pathway containing mapped gene expression values for  $\Delta fhs$  *S. pneumoniae* relative to WT in human serum.** Gene expression values were mapped to gradient colour scale based on the  $\log_2$  fold change expression values of  $\Delta fhs$  compared to WT. Genes highlighted in red are upregulated in this pathway. The intensity of the colour corresponds to fold change values as referenced on the legend.



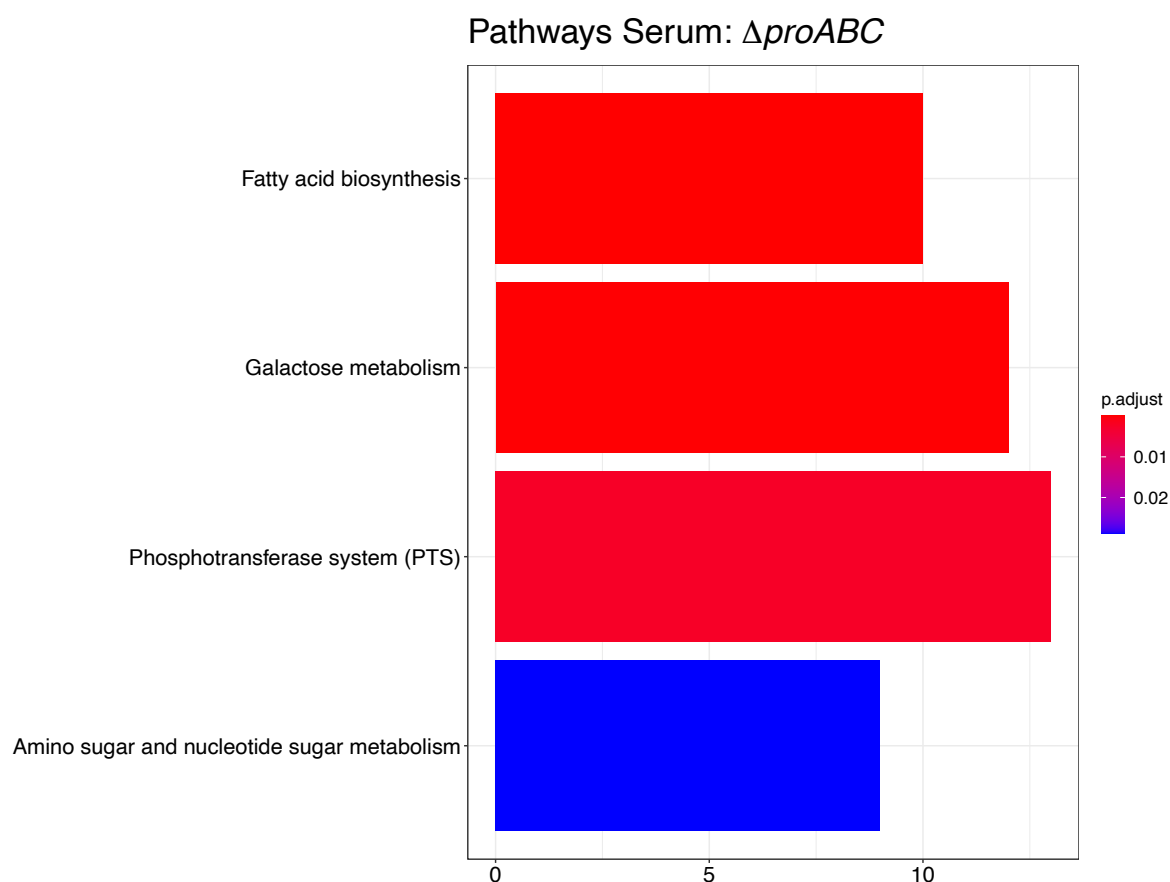
**Figure 5.11: Visualisation of the quorum sensing (competence) KEGG pathway containing mapped gene expression values for  $\Delta fhs$  *S. pneumoniae* relative to the WT in human serum.** Gene expression values were mapped to gradient colour scale based on the  $\log_2$  fold change expression values of  $\Delta fhs$  compared to WT. Genes highlighted in red are up regulated in this pathway. The intensity of the colour corresponds to fold change values as referenced on the legend.



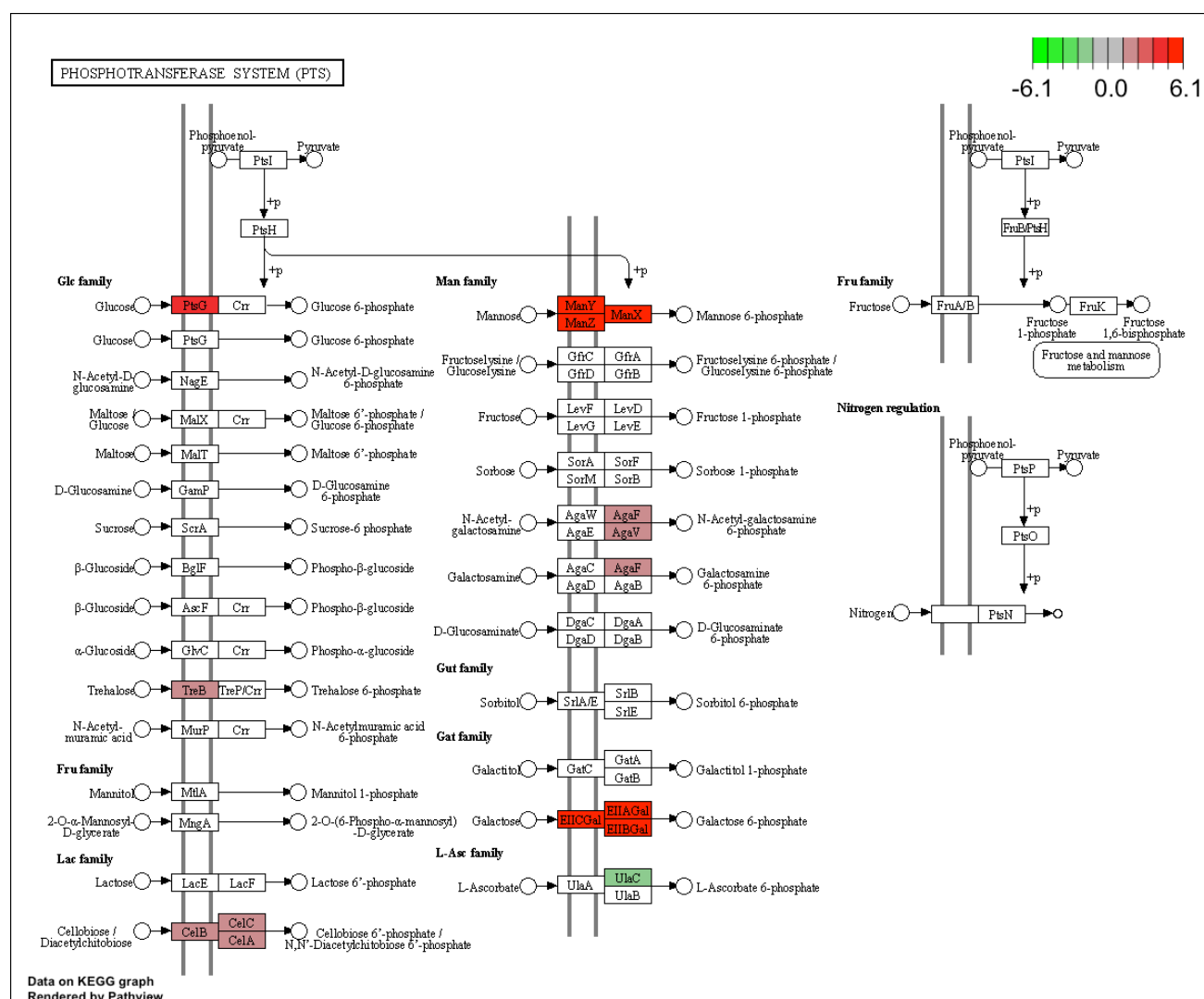
**Figure 5.12: Visualisation of the biosynthesis of secondary metabolites KEGG pathway highlighting the enriched pathways of  $\Delta fhs$  *S. pneumoniae* relative to WT in human serum.** Differentially expressed genes were superimposed on the KEGG metabolic map. The colours indicate the differentially enriched pathways by  $\Delta fhs$  *S. pneumoniae* relative to the WT. The colour of the pathway names corresponds the respective pathways pathway colours. Background pathways in grey were non enriched.

### 5.3.11.2 Biological pathways enriched by $\Delta proABC$ *S. pneumoniae* in human serum

The phosphotransferase system (PTS) and galactose metabolism pathways were significantly enriched by upregulated genes of  $\Delta proABC$  *S. pneumoniae* along with amino sugar and nucleotide sugar metabolism pathways (**Figure 5.13**). **Figure 5.14** and **Figure 5.15** respectively show the  $\Delta proABC$  *S. pneumoniae* gene expression results superimposed on the KEGG PTS and galactose metabolism pathways.

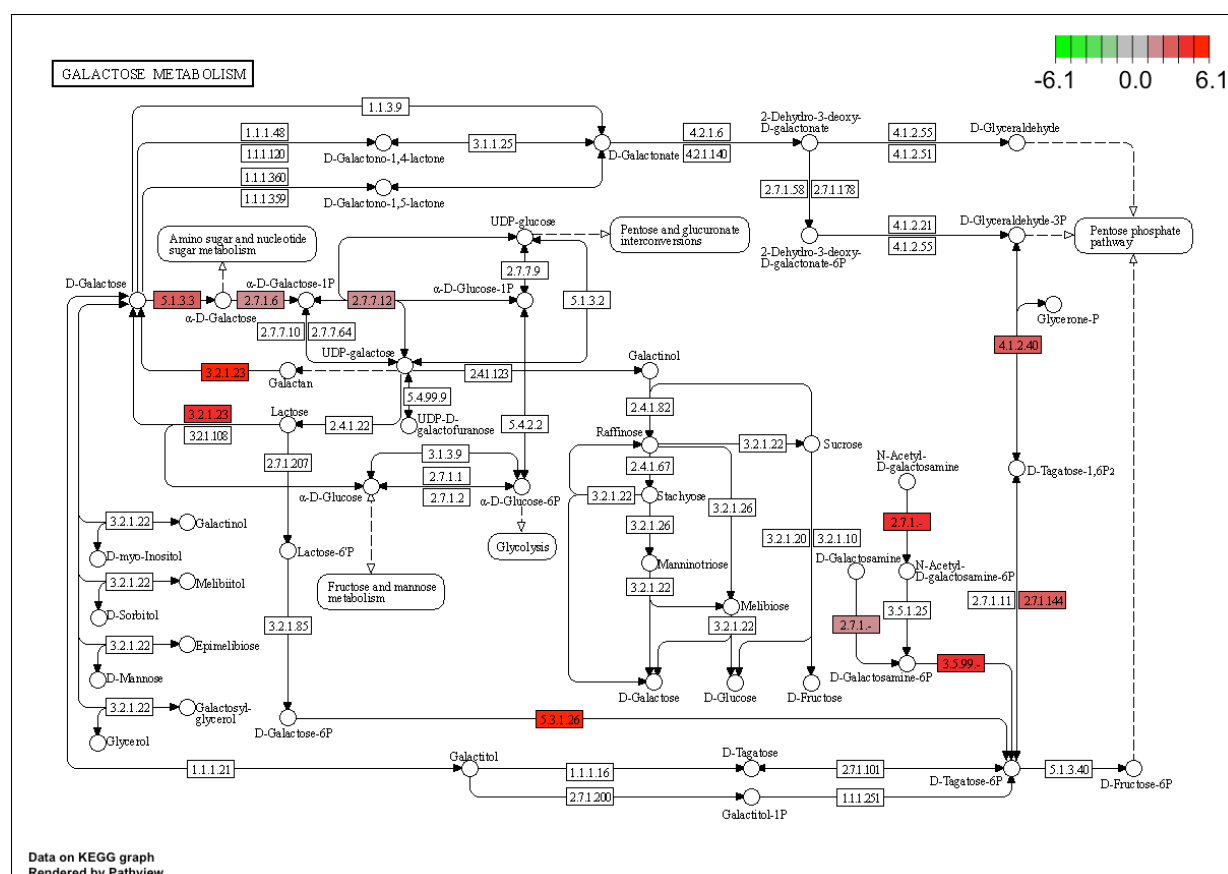


**Figure 5.13: Pathway enrichment analysis of up-regulated genes by  $\Delta proABC$  *S. pneumoniae* during cultured in human serum.** Pathway enrichment was performed by over representation analysis. Pathways with adjusted p-value <0.05 were considered to be significantly enriched. All pathways shown in the plot were significantly enriched with the length of the bar corresponding to the number of DEGs in respective pathways. The colour of the bars denotes level of statistical significance based on adjusted p-value with red being most significantly enriched pathways.



**Figure 5.14: Visualisation of the phosphotransferase system (PTS) KEGG pathway showing expression of significantly regulated genes by  $\Delta$ proABC *S. pneumoniae* during culture in human serum.** Gene expression values were mapped to gradient colour scale based on the  $\log_2$  fold change expression values of  $\Delta$ proABC relative to the WT strain. Genes highlighted in red were up-regulated in and green highlighted were down-regulated. The intensity of the colour corresponds to fold change values referenced on the legend.





**Figure 5.15: Visualisation of the galactose metabolism KEGG pathway showing expression of significantly regulated genes by  $\Delta proABC$  *S. pneumoniae* during culture in human serum.** Gene expression values were mapped to gradient colour scale based on the  $\log_2$  fold change expression values of  $\Delta proABC$  relative to the WT strain. Genes highlighted in red were up regulated in this pathway. The intensity of the colour corresponds to fold change values referenced on the legend.

### 5.3.12 Gene co-expression and module identification

Transcriptomic adaptation involves a coordinated expression of groups of genes involved in specific biological processes. Identification of these sets of genes or modules can provide important insights into underlying biological mechanisms (Eisen *et al.* 1998). Gene co-expression analyses have been used to discover novel gene connections and interdependencies in pneumococcal research (Jamalkandi *et al.* 2020; Aprianto *et al.* 2018). As part of methods development for transcriptomics analyses in this thesis, unsupervised machine learning methods were employed to identify modules of co-expressed genes among the DEGs of  $\Delta fhs$  and  $\Delta proABC$  *S. pneumoniae* cultured in serum. Gene enrichment analysis was then performed for each module by over representation analysis (ORA).

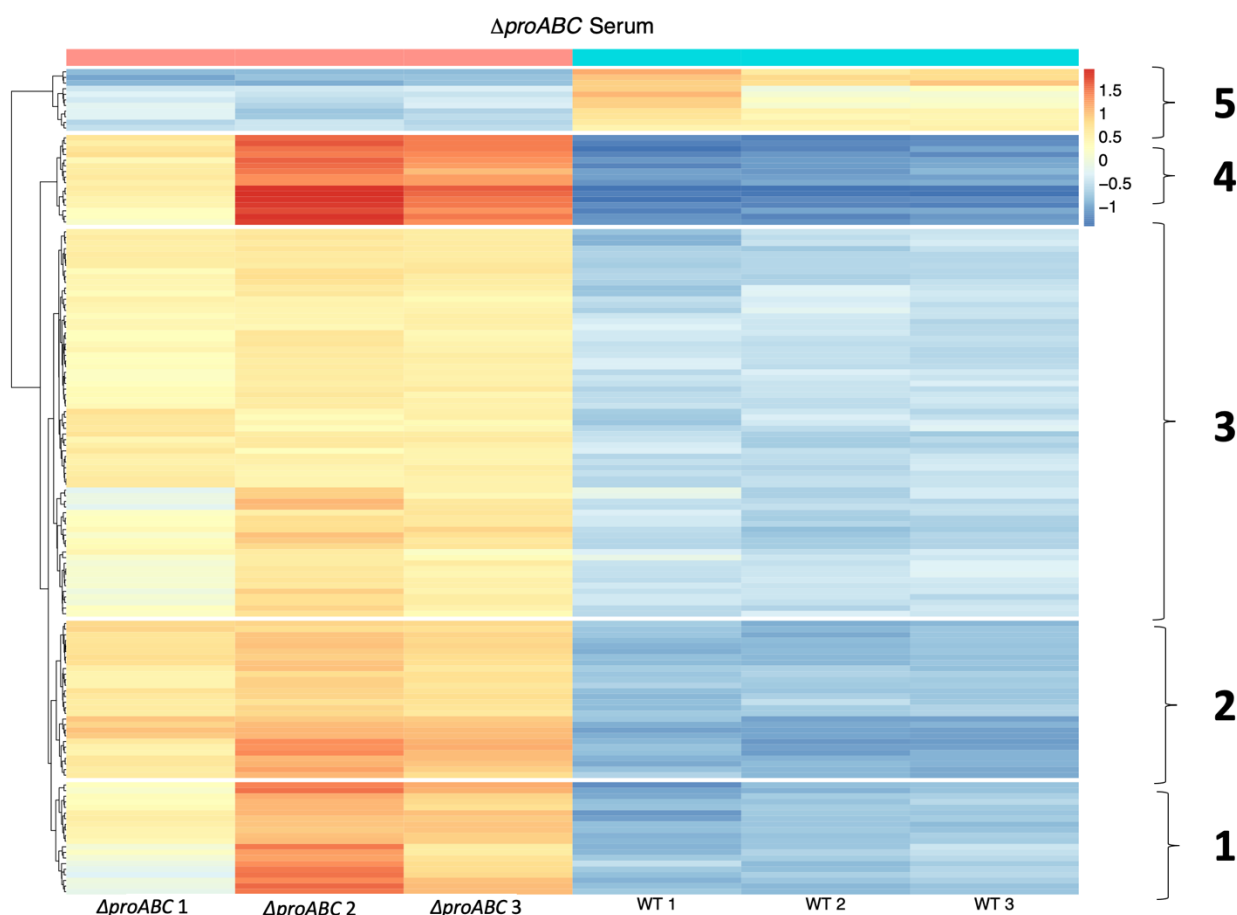
#### 5.3.12.1 Gene modules identified by hierarchical clustering for $\Delta fhs$ *S. pneumoniae*

Gene expression values of DEGs were averaged across all samples and gene co-expression was determined by unsupervised hierarchical clustering using a grouping cut-off (k-mean) predetermined by NbClust. NbClust identifies the optimal number of k-means by selecting the best of 26 different indices (Charrad *et al.* 2014; Eisen *et al.* 1998). The NbClust algorithm proposed k-means of two for the 128 differentially regulated genes of  $\Delta fhs$  *S. pneumoniae* relative to the WT. The analysis did not provide additional insight as DEGs were segregated into up- and down-regulated genes, matching the original results using a significance cut-off.

#### 5.3.12.2 Gene modules identified by hierarchical clustering for $\Delta proABC$ *S. pneumoniae*

Five co-expressed gene clusters were proposed by NbClust for the 144  $\Delta proABC$  *S. pneumoniae* DEGs by hierarchical clustering (**Figure 5.16**). Pathway analyses of the gene modules identified significantly enriched biological pathways in 3 of the 5 clusters (**Table 5.5**). Pathway analysis of the three gene modules replicated results of the analysis of upregulated genes (**Figure 5.13**) but also identified additionally enriched biological pathways. The pathway analysis of results of the three modules that were enriched for at least one biological pathway is shown in **Table 5.5**. Fructose and mannose metabolism, pyruvate metabolism, purine metabolism, and propanoate metabolism were the additionally enriched pathways identified for  $\Delta proABC$  *S. pneumoniae* by applying

hierarchical clustering methods (**Table 5.5**). These results support the hypothesis that deletion of *proABC* in *S. pneumoniae* requires the bacteria to compensate for its functions by upregulating alternative metabolic pathways, and mainly pathways involved with carbohydrate and amino acids biosynthesis.



**Figure 5.16: Gene expression heatmap showing the 5 co-expression gene modules identified by hierarchical clustering for  $\Delta proABC$  *S. pneumoniae* cultured in human serum.** Gene expression values were averaged across all samples and gene co-regulation was determined with hierarchical clustering with a grouping cut-off (k-means) determined by NbClust (Charrad *et al.* 2014). The annotated numbers (1 - 5) on vertical axis represent co-expressed gene modules. Gene expression was averaged by samples ( $n = 6$ ) and displayed as z-scores across each row. The change of colour from blue to red represents the change in z-score of fold change from down- to up-regulated.

**Table 5.5:** Pathway enrichment analyses of the five gene modules identified by hierarchical clustering of 144 DEGs of  $\Delta proABC$  *S. pneumoniae* (Abu-Jamous and Kelly 2018). Pathway enrichment was performed by over representation analysis. Pathways with adjusted p-value <0.05 were considered to be significantly enriched. Cluster 4 (n = 15 genes) and Cluster 5 (n = 11 genes) were not enriched for any biological pathways.

Cluster 1 (n = 20 genes)						
ID	Description	GeneRatio	BgRatio	p.adjust	qvalue	Count
snb02060	Phosphotransferase system (PTS)	9/13	49/733	5.23E-08	2.360E-08	9
snb00520	Amino sugar and nucleotide sugar metabolism	5/13	36/733	0.0007	0.0003	5
snb00051	Fructose and mannose metabolism	3/13	23/733	0.0146	0.0066	3
snb00052	Galactose metabolism	3/13	32/733	0.028	0.0126	3
Cluster 2 (n = 33 genes)						
snb00061	Fatty acid biosynthesis	9/21	12/733	3.66E-12	2.2333E-12	9
snb00640	Propanoate metabolism	4/21	10/733	0.0005	0.00036	4
snb00620	Pyruvate metabolism	4/21	21/733	0.0103	0.00630	4
snb00052	Galactose metabolism	4/21	32/733	0.0336	0.02050	4
Cluster 3 (n = 65 genes)						
snb00230	Purine metabolism	7/30	36/733	0.0071	0.00644	7

### 5.3.12.3 Gene modules identification using Clust algorithm

Next, we employed Clust, a recently developed algorithm for identifying biologically relevant co-expression gene modules from transcriptomic data (Abu-Jamous and Kelly 2018). In contrast to hierarchical clustering, Clust does not force all genes into modules and uses a statistical model to determine only genes that are consistently co-expressed (well correlated) for each module (Abu-Jamous and Kelly 2018). Clust has been shown to

outperform the most commonly used clustering methods for co-expression module identification (Abu-Jamous and Kelly 2018).

#### 5.3.12.4 Gene modules identified by Clust for *Δfhs S. pneumoniae*

The 128 DEGs of *Δfhs S. pneumoniae* in serum were categorized by Clust to consist of two clusters; Cluster A contained 57 genes and Cluster B contained 36 genes. Thirty-two DEGs did not follow a co-expression pattern and therefore did not belong to either Clust gene module. Pathway analysis of the two Clust modules showed significant enrichment for biological pathways by *Δfhs S. pneumoniae* in Cluster A but not in Cluster B. The pathways enriched in Cluster A (**Table 5.6**) recapitulated the pathways identified by applying ORA to all upregulated DEGs for *Δfhs S. pneumoniae* (**Figure 5.9**), suggesting that the subset of DEGs in Cluster A, which are co-ordinately co-expressed, may be driving biological response of *Δfhs S. pneumoniae* during culture in serum.

**Table 5.6:** Pathway enrichment analysis of Cluster A (n= 47 genes annotated in KEGG) module identified by Clust algorithm for *Δfhs S. pneumoniae* cultured in human serum (Abu-Jamous and Kelly 2018). Pathway enrichment was performed by over representation analysis. Pathways with adjusted p-value <0.05 were considered to be significantly enriched. The set of genes enriched for these KEGG pathways were co-expressed by *Δfhs S. pneumoniae* based on Clust analyses (Abu-Jamous and Kelly 2018). Cluster B (n = 36 genes annotated in KEGG) was not enriched for any biological pathways.

ID	Description	GeneRatio	BgRatio	p.adjust	qvalue
snb00061	Fatty acid biosynthesis	9/22	12/733	5.53E-12	2.40E-12
snb00640	Propanoate metabolism	4/22	10/733	6.47E-04	2.81E-04
snb00770	Pantothenate and CoA biosynthesis	4/22	11/733	7.48E-04	3.24E-04
snb01110	Biosynthesis of secondary metabolites	12/22	159/733	2.13E-03	9.22E-04
snb00620	Pyruvate metabolism	4/22	21/733	7.40E-03	3.21E-03
snb00230	Purine metabolism	4/22	36/733	4.62E-02	2.00E-02

### 5.3.12.5 Gene modules identified by Clust for $\Delta$ proABC *S. pneumoniae*

Applying the Clust algorithm to the 144 differentially regulated genes by  $\Delta$ proABC *S. pneumoniae* during culture in serum identified two co-expression gene modules; Cluster A consisted of 54 genes and Cluster B consisted of 51 genes. Thirty-seven genes were not co-regulated and excluded from both clusters. Pathway analysis indicated that Cluster B was not enriched for any biological pathway. The subset of DEGs in cluster A were enriched for the same biological pathways (**Table 5.7**) as considering all DEGs by  $\Delta$ proABC *S. pneumoniae* (**Figure 5.13**), validating the earlier results and indicating that the biological response of the  $\Delta$ proABC *S. pneumoniae* may be mainly driven by a subset of core DEGs.

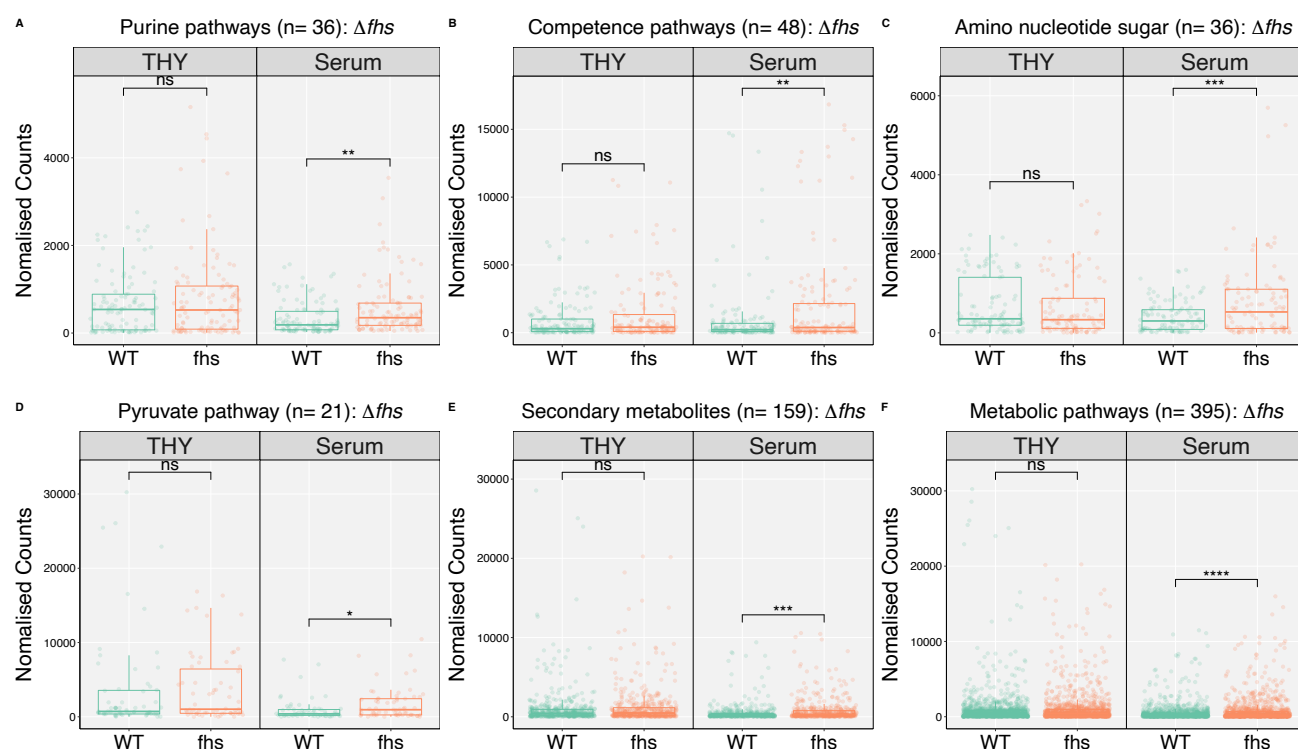
**Table 5.7:** Pathway enrichment analyses of cluster A (n= 47 genes annotated in KEGG) module identified by Clust algorithm for  $\Delta$ proABC *S. pneumoniae* cultured in human serum (Abu-Jamous and Kelly 2018). Pathway enrichment was performed by over representation analysis. Pathways with adjusted p-value <0.05 were considered to be significantly enriched. The set of genes that are enriched for these KEGG pathways were co-expressed by  $\Delta$ proABC *S. pneumoniae* based on Clust analyses (Abu-Jamous and Kelly 2018). Cluster B (n= 50 genes annotated in KEGG) was not enriched for any biological pathways.

ID	Description	GeneRatio	BgRatio	p.adjust	qvalue	Count
snb02060	Phosphotransferase system (PTS)	9/25	49/733	0.00023	0.00013	9
snb00052	Galactose metabolism	7/25	32/733	0.00042	0.00024	7
snb00520	Amino sugar and nucleotide sugar metabolism	5/25	36/733	0.02521	0.01459	5
snb00061	Fatty acid biosynthesis	3/25	12/733	0.02521	0.01459	3

### 5.3.13 Genes in selected pathways were up regulated in serum but not THY for mutant strains.

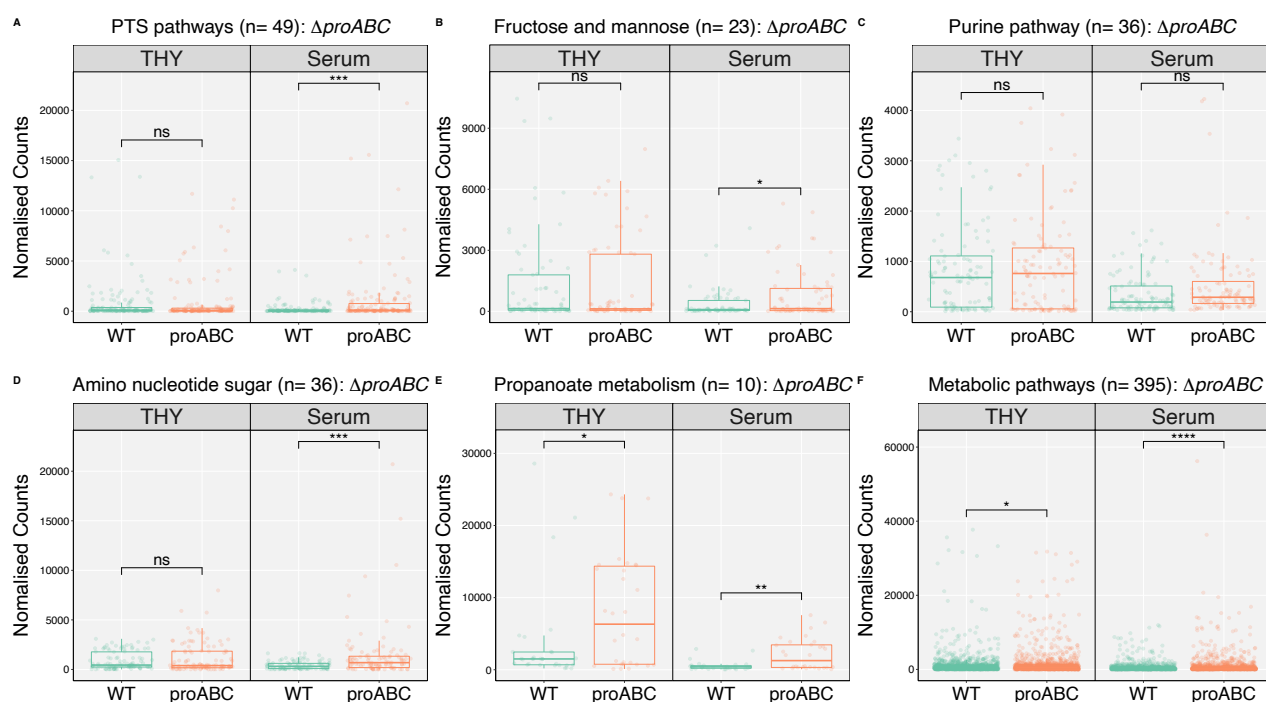
ORA tests whether known biological functions are disproportionately over-represented (enriched) given a list of genes based on hypergeometric distribution (Goeman and Buhlmann 2007). ORA does not account for the magnitude of differential expression from a list of DEGs and the number of genes in a list are based on arbitrary thresholds used to categorise DEGs. Given these limitations of ORA analysis, and because a relatively conservative cut-off was used to categorise DEGs (LFC >1.5 and FDR <0.05), the trend in expression of all the genes that makeup selected pathways was evaluated. DESeq2 normalised gene expression values of  $\Delta fhs$  and  $\Delta proABC$  *S. pneumoniae* were compared to WT's expression values grown in either THY or serum.

The expression of all the genes in the selected pathways were found to be mostly upregulated by  $\Delta fhs$  (**Figure 5.17**) and  $\Delta proABC$  (**Figure 5.18**) *S. pneumoniae* during culture in serum compared to WT. In contrast, there was mostly non-significant differences in the expression of the genes between the mutants and the WT in THY (**Figure 5.17** and **Figure 5.18**). For example, some amino acid biosynthesis pathway genes like *ilvACN* were upregulated by  $\Delta fhs$  *S. pneumoniae* relative to the WT in THY (**Figure 5.5B**), but when the expression of the entire geneset in the amino acid biosynthesis pathway were considered, the pathway was collectively upregulated in serum but not in THY (**Figure 5.17**). This result supports the stress response mechanism of the mutants triggering significant metabolic rewiring during culture in serum in contrast to a more stable metabolic behaviour when grown in enriched media THY. The biosynthesis of secondary metabolites pathway for example, was significant enriched by ORA analysis for  $\Delta fhs$  *S. pneumoniae* grown in both serum and THY. However, when pathway geneset analysis was conducted, there was no difference in the gene expression of this pathway between the  $\Delta fhs$  *S. pneumoniae* and the WT strains in THY (**Figure 5.16E**), suggesting only transient upregulation of some genes in the pathway occurs during growth in THY. Thus, the analysis of evaluating the expression of all of the genes of the pathways therefore provided additional biological insight. ORA analysis was important as a first step to informing pathways to select and interrogate further by this method.



**Figure 5.17: Comparison of the expression of all genes in selected pathways by  $\Delta fhs$  and WT *S. pneumoniae* during growth in THY and human serum.** DESeq2 normalised expressions of the genes within selected pathways (A – F) by  $\Delta fhs$  (n= 3 biological replicates) and WT (n= 3 biological replicates) *S. pneumoniae*. Gene expressions were plotted to identify group differences between the  $\Delta fhs$  and WT strains in THY and serum. The total number of genes from each pathway is stated on individual graphs. Group comparisons were done by unpaired t-test and reported p-values were corrected for multiple testing by Bonferroni method. Green colour dots are normalised expression values for the WT and orange colour dots are the corresponding expression values for  $\Delta fhs$  *S. pneumoniae*.





**Figure 5.18: Comparison of the expression of all genes in selected pathways by  $\Delta proABC$  and WT *S. pneumoniae* during growth in THY and human serum.**

DESeq2 normalise transcript counts of genes within selected pathways (A – F) by  $\Delta proABC$  (n = 3) and WT (n = 3) *S. pneumoniae*. Gene expressions were plotted to identify group differences between  $\Delta proABC$  and WT strains in THY and human serum. The total number of genes in each KEGG pathway is stated on individual graphs. Group comparisons were done by unpaired t-test and reported p-values were corrected for multiple tests by Bonferroni method. Green colour dots are normalised expression values for the WT and orange colour dots are the corresponding expression values for the  $\Delta proABC$  *S. pneumoniae*.

## 5.4 Discussion

*S. pneumoniae* mutants  $\Delta fhs/piaA$  and  $\Delta proABC/piaA$  colonise the nasopharynx of mice similar to WT *S. pneumoniae* serotype 6B but are attenuated in their ability to cause invasive disease (Ramos-Sevillano, Ercoli, Felgner, *et al.* 2021). This provides potential candidate strains to use on healthy human subjects to induce mucosal and systemic immunity (Ramos-Sevillano, Ercoli, Felgner, *et al.* 2021). To expand on these results and to gain insights into the biological mechanisms of attenuation, we conducted an *in vitro* transcriptomic experiment to investigate stress response of pneumococcal knockout strains  $\Delta fhs$  and  $\Delta proABC$  in enriched growth medium THY and nutrient limited medium serum. Growth of bacterial pathogens in nutrient limited environments can identify potential virulence proteins that help the bacteria to overcome physiological stresses during infection (Nanduri *et al.* 2008). A bioinformatics pipeline for processing RNA-Seq data of *S. pneumoniae* and identify differentially expressed genes and biological pathways was developed that can be adapted to compare gene expression and regulation of *S. pneumoniae* serotypes 1 and serotype 6B exposed to different temperatures and iron concentration levels. Through this exercise, we have established an efficient and streamlined workflow for data quality control, mapping, and differential gene expression analysis, utilising open-source bioinformatics software and developing original scripts for RNA-Seq data visualisation. A significant amount of time was spent on trial and error to compare analysis outputs from different software settings and select optimal settings for our purposes.

Culturing  $\Delta fhs$  and  $\Delta proABC$  *S. pneumoniae* in serum induced a significantly broader gene expression profile compared to growth in the enriched media THY. For the genes that were upregulated in THY by both mutant strains, analysis of the expression of the entire gene sets in the respective biological pathways showed no significant difference between the mutant strains and the WT, in contrast to growth in serum (**Figure 5.17** and **Figure 5.18**). THY has high content of glucose, amino acid, and proteins (Todd and Hewitt 1932; Sanchez-Rosario and Johnson 2021), and since *fhs* and *proABC* are involved in the synthesis of several amino-acids (Afzal, Shafeeq, and Kuipers 2016; Crowley *et al.* 1997; UniProt 2021; Kanehisa 2000; Belitsky *et al.* 2001), we hypothesise that the mutants were able to incorporate the end products made by proteins encoded by *fhs* and *proABC* from

the exogenous environment when grown in THY media. Consequently, in nutrient limited environment such as serum, the mutant strains adapt by upregulating the expression of alternative biochemical pathways to try and compensate for the loss of *fhs* and *proABC*.

The genes upregulated in serum were largely distinct between the  $\Delta fhs$  and  $\Delta proABC$  strains as might be expected from the differences in the biological mechanism of action between the mutants. However, Fatty-acid synthesis (FAS) genes were the most significantly upregulated genes in both mutants irrespective of growth environment (**Figures 5.8**). Fatty acids are essential to maintain cell viability and are synthesised in Gram positive organisms through the type II fatty acid synthesis (FASII) pathway (Delekta *et al.* 2018). The expression of this pathway should be downregulated in lipid rich media such as human serum (Brinster *et al.* 2009; Yao and Rock 2017). Because we could not identify plausible biological links between *fhs/proABC* deletion and upregulation of FAS pathway in *S. pneumoniae*, we suspect that upregulation of FAS genes is either a general response to physiological stress or an off-target effect of the gene deletion process. The methods used to generate gene knockdowns in clinical *S. pneumoniae* strains are several decades old and are based on artificially inducing transformation of bacteria with competence-stimulating peptides (Havarstein, Coomaraswamy, and Morrison 1995), followed by confirmation of gene deletions by PCR and southern blotting, and more recently by whole genome sequencing. Bacterial genomes are densely packed, and many genes encode 'moonlighting' proteins with multiple functions that are not evident from genome analysis (Aprianto *et al.* 2018). It is possible that whole gene deletions affect regulatory RNAs such as small RNAs, which are involved in modulating bacterial transcription, translation, mRNA stability, and sensing, thereby altering bacterial gene regulation (Waters and Storz 2009). Therefore, while bacterial genome integrity and genetic architecture may not be affected by deletion of single genes, the influence of such deletions on gene regulation may be less targeted than previously envisioned.

Significantly, culture of both the  $\Delta fhs$  and  $\Delta proABC$  *S. pneumoniae* strains in serum upregulated expression of well-known pneumococcal virulence genes *ply*, *psaA*, and *nanA*. These virulence genes upregulated by the mutants are implicated in *S. pneumoniae* adherence and colonisation. *psaA* plays a critical role in regulation of gene expression of

*S. pneumoniae* in response to oxidative stress response and intracellular redox homeostasis (Tseng *et al.* 2002) and is also implicated to promote bacterial adhesion to host cells (Rajam *et al.* 2008). *nanA* enhances adherence of *S. pneumoniae* to the mucosal surfaces and is also involved in biofilm formation (Parker *et al.* 2009; Manco *et al.* 2006). Pneumolysin is a bacterial toxin lyses cells (Shak *et al.* 2013), that promotes biofilm formation, plays a role in colonisation and transmission of *S. pneumoniae* in the nasopharynx (Zafar *et al.* 2017; Hotomi *et al.* 2016), and is involved in regulating the complement system. Under nutrient poor conditions, *S. pneumoniae* is known to increase expression of *ply* for survival (Zafar *et al.* 2017). Preliminary *in vitro* analysis by Dr Caroline Weight at UCL suggests that both  $\Delta fhs$  and  $\Delta proABC$  *S. pneumoniae* adhere less during culture in epithelial cells compared to WT strains (Unpublished). However, the mutants had an elevated microinvasion index which means that if they adhered, they were more likely to be internalised compared to the WT (Weight *et al.* 2018). In this context, the upregulation of virulence genes *nanA* and *psaA* genes, both of which are involved in enhancing pneumococcal attachment to epithelial cells (Tseng *et al.* 2002; Berry and Paton 1996; Manco *et al.* 2006; Parker *et al.* 2009), may represent an attempt by the mutant strains to increase bacterial adherence capacity. *psaA* also plays a central role in bacterial response to oxidative stress and it is possible that the significant rewiring of the bacterial metabolism due to the gene deletions may be associated with change in expression of *psaA* (Berry and Paton 1996).

Functional analysis identified bacterial pathways for both mutants which were dominated by metabolic pathways genes. *fhs* and *proABC* are putatively involved with biosynthesis and metabolic processes (Afzal, Shafeeq, and Kuipers 2016; Crowley *et al.* 1997; Ramos-Sevillano, Ercoli, Felgner, *et al.* 2021), and the deletion of these genes showed significant metabolic rewiring of the transcriptome of both mutant strains compared to the WT. Consistent with the hypothesis that gene expression of  $\Delta fhs$  strain will represent metabolic compensation for loss of *fhs*, the purine metabolism pathway was significantly enriched in  $\Delta fhs$  *S. pneumoniae* in serum. The quorum sensing (competence) pathway was also distinctly enriched in the  $\Delta fhs$  strain. Induction of competence genes is purported as a general stress response mechanism of *S. pneumoniae* which can be exacerbated by cell-to-cell contact (Prudhomme *et al.* 2016; Claverys, Prudhomme, and Martin 2006). The

upregulation of virulence genes *psaA*, *ply*, and *nanA*, all involved in enhancing pneumococcal attachment to epithelial surfaces, together with upregulation of competence regulon in a nutrient limited environment may explain the maintained ability of the  $\Delta fhs$  *S. pneumoniae* to colonise the nasopharynx in murine models (Ramos-Sevillano, Ercoli, Felgner, *et al.* 2021) which may include biofilm formation. Consistent with biofilm formation for adaptation of *S. pneumoniae* in the nasopharynx,  $\Delta fhs$  *S. pneumoniae* upregulated expression of competence genes and *nanA* in addition to non-significant expression of capsular genes (Marks, Reddinger, and Hakansson 2012; Parker *et al.* 2009). The functions of  $\Delta proABC$  proteins have not been formally proven but they are very likely to encode enzymes required for the synthesis of L-proline from L-glutamate, and this is supported by the partial complementation of growth defects of the  $\Delta proABC$  *S. pneumoniae* strain in defined media by addition of proline (Ramos-Sevillano Unpublished data) (UniProt 2021; Kanehisa 2000; Belitsky *et al.* 2001). However, based on our functional analysis, *proABC* also has a key role in *S. pneumoniae* biosynthesis and utilisation of carbohydrates.  $\Delta proABC$  *S. pneumoniae* was almost exclusively enriched for phosphotransferase system (PTS) and several carbohydrate metabolism pathways (e.g., galactose metabolism, fructose, and mannose metabolism, etc). *S. pneumoniae* depends on carbohydrate metabolism for energy and growth (Paixao, Caldas, *et al.* 2015) and a large part of its genome is dedicated to the uptake and metabolism of carbohydrates (Tettelin *et al.* 2001). *S. pneumoniae* is known to metabolise at least thirty-two different carbohydrates (Tettelin *et al.* 2001; Bidossi *et al.* 2012), which is critical in its ability to adhere to and colonise the nasopharynx (King 2010), produce its capsule, and for virulence (Paixao, Caldas, *et al.* 2015; Iyer and Camilli 2007; King 2010).

In conclusion, we have developed an *S. pneumoniae* RNA-Seq pipeline that highlights the broad metabolic adaptation of  $\Delta fhs$  and  $\Delta proABC$  *S. pneumoniae* in a stress environment. During culture in serum stress, both mutant strains upregulated expression of pneumococcal virulence genes and several metabolic pathways, a classical response mechanism of *S. pneumoniae* to stress. While  $\Delta fhs$  strain mainly upregulated competence and secondary metabolic pathways such as purine metabolism, the  $\Delta proABC$  strain was biased towards upregulation of bacterial pathways involved in amino acid and carbohydrate metabolism and utilisation. Taken together, our data indicates an important

role for pneumococcal metabolism in driving *S. pneumoniae* virulence and identifies biological pathways that may be involved in the adaptation of  $\Delta fhs$  and  $\Delta proABC$  *S. pneumoniae*. The mutants are currently being tested in the experimental human challenge model in Liverpool with the aim of validating *in vivo* results observed in murine studies. Ongoing work is focused on identifying biological linkage between upregulated genesets in the bacterial strains and the epithelial transcriptomic response.

## Chapter 6 : Transcriptomic analysis of the adaptation of *S. pneumoniae* under Iron-deplete and Iron- replete conditions

## 6.1 Introduction

Trace transition metals are one of the most essential elements for all living organisms and act as co-factors in many biological processes (Turner *et al.* 2017). Iron is an essential nutrient for many basic cellular and metabolic processes including respiratory pathways for both the host and the microorganism (Schaible and Kaufmann 2004). Pathogenic bacteria rely on acquisition of essential metal ions such as iron, zinc, and manganese for growth and survival within the host. Due to the importance of iron availability for microbial growth and virulence, including antimicrobial defence, the host has evolved strategies to significantly sequester the availability of iron and other nutrition metals ions (e.g., Zn, Mn) to control infection (Schaible and Kaufmann 2004). In addition, the host releases harmful metals such as copper (Cu) to the site of infection to exert an antimicrobial function (Hood and Skaar 2012; Djoko *et al.* 2015). This concept of host-mediated starvation of beneficial metals ions and release of toxic metals against microbes during infection is referred to as “nutritional immunity” (Kehl-Fie and Skaar 2010; Hood and Skaar 2012).

Our understanding of nutritional immunity is supported by both epidemiological and experimental data that show a direct correlation between host iron uptake and increased risk of infection. Iron overload due to nutritional heredity (e.g., haemochromatosis), therapeutics (e.g., prophylactic iron supplementation), or refeeding with iron supplementation following malnutrition has been shown to exacerbate various infectious diseases including salmonellosis, tuberculosis, AIDS, and yersinioses (Pietrangelo 2004; Moalem, Weinberg, and Percy 2004; Murray *et al.* 1978; Weinberg 1999; Cunningham-Rundles *et al.* 2000). Iron concentrations of  $10^{-6}$  to  $10^{-7}$ M are required by most microbes for critical metabolic processes related to microbial replication, glycolysis, electron transport, defence against ROI, and DNA synthesis (Schaible and Kaufmann 2004). However, the host can substantially reduce this concentration to  $10^{-18}$ M through iron sequestration (Sheldon, Laakso, and Heinrichs 2016). In the host, iron is available highly bound to specific protein structures with about 80% being haem-bound (e.g., haemoglobin and myoglobin) and the rest bound to other proteins such as transferrin, lactoferrin, and ferritin (Schaible and Kaufmann 2004; van Beek *et al.* 2020). Most microorganisms have evolved elaborate mechanisms to access iron in the nutrient limited



environment *in vivo* through specialised iron uptake systems that compete for iron in the extracellular matrix (Schaible and Kaufmann 2004). To improve access to the host iron pool, some bacterial pathogens such as *Staphylococcus aureus* and *Neisseria meningitidis* encode haem-transport systems (HTS) to preferentially target the more abundant haem-complexed iron rather than the transferrin bound iron (transferrin accounts for only 1% of the iron pool) (Schaible and Kaufmann 2004; Skaar *et al.* 2004; Hood and Skaar 2012). Some bacteria such as *Escherichia coli* and *Staphylococcus aureus* also secrete and import siderophores, small high affinity iron binding molecules, that load and uptake iron from lactoferrin and other host iron-binding proteins (Wilson *et al.* 2016). This constant battle between the host and microorganisms for available iron can be a determinant of infection progression (Wilson *et al.* 2016).

In contrast to Gram-negative pathogens (e.g., *Salmonella* and *Neisseria*) and some Gram-positive bacteria (e.g., *S. aureus*), iron acquisition and sensing mechanisms by *S. pneumoniae* have been less well defined. Three iron uptake ABC transporters systems *PiaABCD* (originally named *Pit2ABCD*), *PiuBCDA* (originally named *Pit1ABCD*), and *PitABCD* were the first to be described in *S. pneumoniae* by Brown and colleagues in early 2000s (Brown, Gilliland, and Holden 2001; Brown *et al.* 2002; Brown and Holden 2002). However, the mechanism of iron acquisition by these proteins has not been fully elucidated. Double mutant deletions of *piaA* and *piuB* were shown to be unable to utilise haemoglobin as iron source and are attenuated for virulence in murine infection models. However, single gene mutants do not exhibit this phenotype suggesting that the iron transporters work independently and that one is able to compensate for the loss of the other (Brown, Gilliland, and Holden 2001). The involvement of *pitA* in iron acquisition is demonstrated through enhanced growth impairment when a *pitA* mutation was added to a strain already carrying mutations in *piaA* and *piuA* (Brown *et al.* 2002). A single deletion of *pitA* has no growth or virulence effect suggesting that *pitABCD* is not as important for *S. pneumoniae* virulence as *piaABCD* and *piuBCDA* (Brown *et al.* 2002). Using proteomics, Yang *et al.*, recently described a putative novel iron binding protein (*SPD\_1609*) in a D39 (serotype 2) strain, which was suggested to functions similar to *pitA* (Yang *et al.* 2016). Attempts to generate a *pitA/piaA/piuA/SPD\_1609* tetra mutant strain did not result in viable bacterial colonies with the authors suggesting complete loss of bacterial iron uptake ability as a

possible explanation (Yang *et al.* 2016). The same group has also recently identified *SPD\_1590*, previously annotated as a general response protein (*gls24*), as an *S. pneumoniae* iron transporter that may have a similar function to *piaA* (Miao *et al.* 2018; Tettelin *et al.* 2001). Unlike *pitA*, *piaA* and *piaA* are highly conserved in *S. pneumoniae* and sera from septicemia patients show high antibody reactivity to *piaA* and *piaA* (Whalan *et al.* 2005; Jomaa *et al.* 2006; Whalan *et al.* 2006). Another *S. pneumoniae* protein thought to be important for iron homeostasis is *dpr*, an iron storage protein that also plays a key role in *S. pneumoniae* response to oxidative stress (Pericone *et al.* 2003; Hua *et al.* 2014). The regulators suggested to drive pneumococcal iron transport and homeostasis include *ritR* (Ulijasz *et al.* 2004; Glanville *et al.* 2018) and *psaR* (Ong *et al.* 2013). *ritR* is thought to be the master regulator of iron in *S. pneumoniae* (Ulijasz *et al.* 2004; Glanville *et al.* 2018). **Table 6.1** summarises the *S. pneumoniae* genes previously shown to be involved in transition metals import and export. *S. pneumoniae* does not produce siderophores or iron chelating growth factors during iron limitation (Tai, Lee, and Winter 1993) and evidence of genes encoding siderophore proteins is absent from *in silico* analysis (Brown and Holden 2002; Tettelin *et al.* 2001).

Most of the discovery studies on iron uptake genes in *S. pneumoniae* have been conducted using the laboratory adapted strain D39 (serotype 2). As previously mentioned, the D39 strain which was first isolated in 1914 is the most commonly used strain in *S. pneumoniae* research and has been passaged for more than a century (Griffith 1928; Avery, Macleod, and McCarty 1944; Slager, Aprianto, and Veening 2018). The strain has been used in many seminal discoveries including the work of Avery *et al.*, which identified DNA as the hereditary substance of living organisms. However, laboratory adapted strains like D39 have been shown to have several genomic (including ABC transporter genes) and metabolic differences compared to clinical strains (Hoskins *et al.* 2001; Lanie *et al.* 2007). Extrapolation of findings from D39 to clinical strains may be incomplete and thus it is possible that more bacterial elements such as iron uptake systems are yet to be discovered in *S. pneumoniae* clinical strains.

The advent and accessibility of next-generation high-throughput sequencing methods allow for higher resolution investigation of global gene expression of *S. pneumoniae*

compared to previous iron studies (Croucher and Thomson 2010; van Vliet 2010; Sharma and Vogel 2014; Poulsen and Vinther 2018). This will allow for a more holistic overview of the interplay and connectivity of the entire *S. pneumoniae* gene set in response to iron availability. It is becoming clear that bacterial pathogens adapt to new environments through sensing regulators and fine-tuning gene expression of dedicated pathways based on nutritional conditions to achieve homeostasis or to source critical by-products for survival (Weiser, Ferreira, and Paton 2018; van Opijnen and Camilli 2012). In this chapter, RNA sequencing was used to investigate the adaptation of clinical *S. pneumoniae* serotype 1 (BVJ1JL) and serotype 6B (BHN418) in iron-deplete and iron-replete growth conditions. *S. pneumoniae* serotype 1 is a highly invasive strain that is rarely isolated in healthy individuals and is a common cause of epidemics in sub-Saharan Africa (Brueggemann *et al.* 2004; Weinberger *et al.* 2010; Ruckinger *et al.* 2009; Kwambana-Adams *et al.* 2016). *S. pneumoniae* serotype 6B on the other hand is a common coloniser of the nasopharynx in healthy individuals and a less common cause of invasive pneumococcal disease (IPD) (Brueggemann *et al.* 2004; Weinberger *et al.* 2010). We hypothesise that altering iron availability will elicit an altered transcriptomic response in *S. pneumoniae*, which in turn will elucidate important virulence factor differences between clinical strains of *S. pneumoniae*.

**Table 6.1:** Previously reported *S. pneumoniae* genes involved in transition metal homeostasis

Operon	Gene	Product	Phenotype/function	Refs.
<i>piuBCDA</i> (Iron import)	<i>piuA</i>	Iron lipoprotein receptor	$\Delta piuA/\Delta piaA$ attenuates virulence in murine infection model	(Brown, Gilliland, and Holden 2001; Brown <i>et al.</i> 2001)
	<i>piuB</i>	Transmembrane permease		
	<i>piuC</i>	Transmembrane permease		
	<i>piuD</i>	ATPase component		
<i>piaABCD</i> (Iron import)	<i>piaA</i>	Iron lipoprotein receptor	$\Delta piuA/\Delta piaA$ attenuates virulence in murine infection model	(Brown, Gilliland, and Holden 2001; Brown <i>et al.</i> 2001)
	<i>piaB</i>	Transmembrane permease		
	<i>piaC</i>	Transmembrane permease		
	<i>piaD</i>	ATPase component		
<i>pitADBC</i> (Iron import)	<i>pitA</i>	Iron lipoprotein receptor	Addition of $\Delta pitA$ to $\Delta piuA/\Delta piaA$	(Brown <i>et al.</i> 2002)
	<i>pitB</i>	Transmembrane permease	impairs growth further.	
	<i>pitC</i>	Transmembrane permease		
	<i>pitD</i>	ATPase component		
	<i>ritR</i>	Repressor of iron transport	Regulator of iron homeostasis. $\Delta ritR$ impairs bacterial growth/ not essential for survival	(Ong <i>et al.</i> 2013; Ulijasz <i>et al.</i> 2004)

(Table Continued)

<i>psaABC</i> (Manganese import)	<i>psaA</i>	Manganese ABC transporter substrate-binding lipoprotein	$\Delta$ <i>psaA</i> , $\Delta$ <i>psaB</i> , $\Delta$ <i>psaC</i> significantly attenuated in murine infection models.	(Berry and Paton 1996; Marra <i>et al.</i> 2002; Johnston <i>et al.</i> 2006)
	<i>psaB</i>	Manganese ABC transporter ATP-binding protein	<i>psaR</i> Regulates	
	<i>psaC</i>	Manganese ABC transporter permease	<i>PsaBCA</i> permease/	
	<i>psaR</i>	Manganese-dependent regulator	essential for murine infection, not colonisation.	
<i>copYA/cupA</i> (Copper import)	<i>copy</i>	cop operon regulator	Operon is essential for pneumococcal colonisation.	(van Opijnen and Camilli 2012; Shafeeq <i>et al.</i> 2011)
	<i>copA</i>	P-type ATPase copper exporter		
	<i>cupA</i>	Copper transport protein and potential cupredoxin		
<i>adcAAI/BCR</i> (Zinc import)	<i>adcA</i>	Zinc Importer (Redundant with <i>AdcAll</i> )	$\Delta$ <i>adcAll</i> negatively impact early colonisation of nasopharynx.	(Brown <i>et al.</i> 2016)
	<i>adcAll</i>	Zinc Importer (Redundant with <i>AdcAll</i> )		
	<i>adcB</i>	Transmembrane permease		
	<i>adcC</i>	ATPase		
	<i>adcR</i>	Import Machinery		
		Transcriptional regulator		
	<i>czcD</i>	Zinc efflux system		
	<i>phtABDE</i>	Zinc importer	Facilitate zinc import via <i>AdcAll</i>	(Plumptre <i>et al.</i> 2014)

(Table Continued)

<i>sodA</i>	Manganese-dependent superoxide dismutase	$\Delta$ <i>sodA</i> attenuated virulence in murine models. Essential for early lung infection.	(Yesilkaya <i>et al.</i> 2000)
<i>mntE</i>	Mn <sup>2+</sup> efflux membrane protein	$\Delta$ <i>mntE</i> attenuated virulence in murine infection model.	(Rosch <i>et al.</i> 2009)

---

## 6.2 Methods

### 6.2.1 Bacterial strains and growth media

The clinical strains of serotype 1 (BVJ1JL) and a serotype 6B (BHN418) were used for all experiments in this chapter. The serotype 1 strain (BVJ1JL) was isolated from the nasopharynx of a 9-year-old child in Malawi in 2015 (Betts *et al.* 2021). The BHN418 strain was isolated from the nasopharynx of a child in Sweden between 1997 – 2004 (Browall *et al.* 2014). The BHN418 strain is currently being used to inoculate healthy adults as part of the controlled Experimental Human Pneumococcal Carriage (EHPC) study in Liverpool (Mitsi *et al.* 2020). The original BHN418 strain has been re-sequenced as part of this thesis using long- and short-read sequencing technologies (Chapter 4) and the genome assembly is referred to herein as BHN418UCL. Characterisation of the re-sequenced complete genome assemblies of BHN418UCL and BVJ1JL strains are described in Chapter 4.

The “nose-mimicking condition” (NMC) media described by Aprianto *et al.* (Aprianto *et al.* 2018) was used as the chemically defined media (CDM) for transcriptomic experiments. NMC is a modification of the Sicard’s defined growth medium (Sicard 1964; Aprianto *et al.* 2018) and contains no iron. For comparison, we also used the enriched growth medium Todd-Hewitt broth + Yeast (THY) which is iron replete and is commonly used for optimal growth of fastidious beta-haemolytic *Streptococci* such as *S. pneumoniae* (Todd and Hewitt 1932; Sanchez-Rosario and Johnson 2021). THY has a rich content of transition metal ions with an estimation of 734 -740µg/L iron concentrations (Hoyer *et al.* 2018; Jimenez-Munguia *et al.* 2018). The base THY growth medium was therefore considered

an iron replete condition for this experiment. The concentration of transition metals in NMC has not been measured (Aprianto *et al.* 2018). Although iron was not directly added as a constituent in preparation of NMC, other transition metal ions such as  $\text{MgCl}_2$ ,  $\text{ZnSO}_4$ , and  $\text{CaCl}_2$  constitute part of the media formulation (see **Appendix 3**). Hoyer and colleagues (Hoyer *et al.* 2018) measured the concentration of iron in a CDM based on a backbone of RPMI 1640 medium to be  $190\mu\text{g/L}$ , four-fold lower than in THY. However, the concentration of iron based on Sicard's medium (NMC) may be different from the estimation of Hoyer and colleagues. All growth in this chapter were carried out statically at incubation conditions of  $37^\circ\text{C}$  and 5%  $\text{CO}_2$ .

### 6.2.2 Induction of iron-deplete and iron-replete conditions

The chemical chelating compound 2,2'-bipyridine (BIP) (Sigma) was used to artificially induce iron stress of *S. pneumoniae* cultured in CDM and THY. BIP is a commonly used chelator for experiments investigating effects of iron limitation on bacterial microorganisms (van Opijnen and Camilli 2012; Hoyer *et al.* 2018; Kroger *et al.* 2013; Kroger *et al.* 2018). BIP forms complexes with iron as well as other transition metal ions (e.g., manganese, zinc etc). To induce an iron-replete condition in CDM,  $\text{FeSO}_4$  was added to *S. pneumoniae* culture grown to mid-exponential phase (MEP) in CDM.  $\text{FeSO}_4$  supplementation of growth media has a similar effect to haemin and haemoglobin (Tai, Lee, and Winter 1993; van Opijnen and Camilli 2012), the primary iron source for bacteria in the host. For iron-deplete and iron-replete conditions, BIP and  $\text{FeSO}_4$  were added to a final concentration of 1mM in media respectively. **Table 6.2** summarises the experimental conditions investigated.

### 6.2.3 Effect of iron depletion on *S. pneumoniae* growth

Growth characteristics of BHN418 and BVJ1JL in base media (THY and CDM), iron-deplete (THY and CDM), and iron-replete (CDM only) media were evaluated using Spark plate reader (Tecan). Optical density of cultures was measured every 30 minutes over an 18hr period with 5 seconds double orbital-shaking preceding each reading. Four technical replicates were included for each experiment and data from three independent experiments was combined for analysis. The aims of these growth curve experiments

were (i) To investigate the effect of BIP, iron (Fe), and manganese (Mn) on *S. pneumoniae* growth (ii) To confirm MEP growth points of serotype 1 (BVJ1JL) and serotype 6B (BHN418UCL) in CDM and THY.

#### 6.2.4 Cellular viability of *S. pneumoniae* under Iron-deplete conditions

BIP removes extracellular excess metals by forming water-soluble complexes to prevent them from interacting with other molecules or enzymes (Holyer *et al.* 2002). To ascertain that this biochemical process was not toxic to bacterial viability and cellular proliferation, the serotype 1 strain (BVJ1JL) was used to investigate the number of viable bacterial counts post addition of BIP to bacterial cultures. BVJ1JL bacteria stored in THY and 10% glycerol were defrosted from -80°C and added to freshly prepared CDM or THY in ~1:83 ratio (300µL of bacterial stock added to 25mL of media). The culture was incubated at 37°C and 5% CO<sub>2</sub> up to MEP and then BIP was added to a final concentration of 1mM. Colonies forming units (CFU/ml) were compared at MEP and at 1hr and 4hr post incubation in 1mM BIP.

#### 6.2.5 *S. pneumoniae* transcriptome in iron-deplete and iron-replete conditions

Transcriptomic experiments were designed to assess transcriptomic adaptation to iron availability and minimise gene expression signals associated with normal *S. pneumoniae* growth. We developed a protocol based on that used by the Hinton Group for *Salmonella* transcriptomics (Kroger *et al.* 2013) which minimises the effects of bacteria growth on the transcriptome by using short exposure times (**Figure 6.1**). The two strains of *S. pneumoniae* (BVJ1JL and BHN418) were grown to MEP in THY and CDM with and without BIP (1mM), and for CDM only, with and without FeSO<sub>4</sub> (1mM). The doubling time of *S. pneumoniae* in the CDM has been estimated to be 62 (±7) mins (Chapter 3) with MEP OD<sub>600</sub> 0.25 – 0.32. The doubling time in THY is ~40mins and reaches MEP at OD<sub>600</sub> 0.3 – 0.35. To understand how the transcriptome of *S. pneumoniae* evolves and adapts over time in iron deplete and iron replete conditions, a time course of 10 minutes and 50 minutes post exposure was opted. Four biological replicates were included for each strain and timepoint. To avoid ongoing gene expression and RNA degradation following the completion of the experiment, the Qiagen RNeasy Protect<sup>®</sup> Bacteria Reagent was added to



stabilise RNA according to manufacturer's protocol and stored at -80°C. The conditions that were investigated for both strains are summarised in **Table 6.2**.

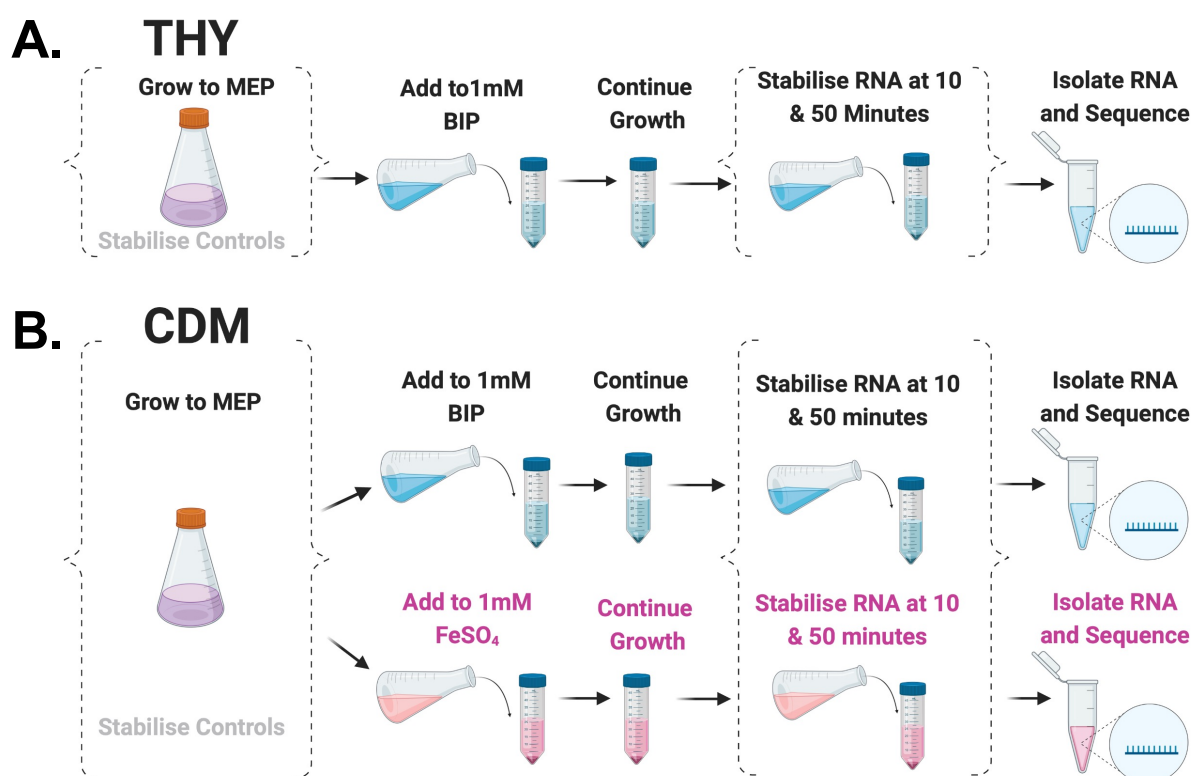
**Table 6.2.** Growth conditions to induce *S. pneumoniae* transcriptome for serotype 1 (BVJ1JL) and serotype 6B (BHN418UCL) strains

Condition	Growth description
THY	Growth in THY to mid exponential phase (MEP)
THY10	Growth in THY to MEP; then addition of 2,2' - dipyridyl (BIP) to final concentration of 1mM and incubated further for 10 mins
THY50	Growth in THY to MEP; then addition of 2,2' - dipyridyl (BIP) (to final concentration of 1mM and incubated further for 50 mins
CDM	Growth in CDM to mid exponential phase MEP
BIP10	Growth in CDM to MEP; then addition of 2,2' - dipyridyl (BIP) to final concentration of 1mM and incubated further for 10 mins
BIP50	Growth in CDM to MEP; then addition of 2,2' - dipyridyl (BIP) to final concentration of 1mM and incubated further for 50 mins
FeSO <sub>4</sub> 10	Growth in CDM to MEP; then addition of FeSO <sub>4</sub> to final concentration of 1mM and incubated further for 10 mins
FeSO <sub>4</sub> 50	Growth in CDM to MEP; then addition of FeSO <sub>4</sub> to final concentration of 1mM and incubated further for 50 mins

### 6.2.6 RNA isolation and quantification

Samples stabilised with RNAprotect® Bacteria Reagent were retrieved from -80°C for RNA isolation. An optimised RNA isolation protocol based on the commercially available MirVana kit (Thermo Fisher) and described in detail in Chapter 2, was used. Total isolated RNA was DNase treated using the TURBO DNase digestion kit (Thermo Fisher) and cleaned up with RNeasy MinElute CleanUp Kit (Qiagen) with a final elution volume of 20µL. RNA quantity and quality were assessed by chip-based capillary electrophoresis (Agilent Bioanalyzer2100), Nanodrop spectrophotometer (Thermo Fisher), and Qubit (Invitrogen)

(using RNA broad range (BR) Assay kit). Ten microliters of triplicates with the best RNA quantification values were submitted to Vertis Biotechnologie AG (Freising, Germany) for RNA sequencing. Twenty-four samples for each strain (forty-eight in total) were shipped in dry ice overnight to Vertis Biotechnologie AG (Freising, Germany).



**Figure 6.1: Experimental design for investigating the transcriptome of *S. pneumoniae* in iron-deplete and iron-replete conditions.** Workflow for investigating iron limitation (1mM BIP) in (A) enriched media Todd-Hewitt broth + Yeast (THY) and (B) a chemically defined media (CDM). Mid-exponential phase (MEP) cultures were exposed to 1mM BIP or 1mM FeSO<sub>4</sub> for 10 minutes and 50 minutes and stabilised with RNA protect reagent (Qiagen). Baseline MEP samples without BIP or FeSO<sub>4</sub> were stabilised and used as transcriptomics reference control. For each condition four biological replicates were included. The dotted bracket lines indicate baseline samples and experiment test samples. RNA was extracted from stabilised samples for RNA sequencing. Figure generated in BioRender (Perkel 2020).

### 6.2.7 Whole transcriptome cDNA

The total RNA samples were first fragmented using ultrasound (4 pulses of 30 s each at 4°C). Then, an oligonucleotide adapter was ligated to the 3' end of the RNA molecules. First-strand cDNA synthesis was performed using M-MLV reverse transcriptase and the 3' adapter as primer. The first-strand cDNA was purified, and the 5' Illumina TruSeq sequencing adapter was ligated to the 3' end of the antisense cDNA. The resulting cDNA was PCR-amplified to about 10-20 ng/μL using a high-fidelity DNA polymerase. The TruSeq barcode sequences which are part of the 5' and 3' TruSeq sequencing adapters were used. The cDNA was purified using the Agencourt AMPure XP kit (Beckman Coulter Genomics) and was analysed by capillary electrophoresis.

### 6.2.8 cDNA from +/- terminator exonuclease (TEX) treatment of RNA (dRNA-Seq)

For differential RNA-Seq (dRNA-Seq) library preparation, the RNA samples of BHN418 (serotype 6B) and BVJ1JL (serotype 1) strains were pooled separately. First, the total RNA samples were fragmented using ultrasound (4 pulses of 30 sec at 4°C). This was followed by a treatment with T4 Polynucleotide Kinase (New England Biolabs). The samples were then split into two halves and one half was subjected to terminator exonuclease treatment (+TEX) and the other half was left untreated (-TEX). The treated RNA samples were poly(A)-tailed using poly(A) polymerase and then, the 5'PPP structures were removed using RNA 5' Polyphosphatase (Epicentre). Afterwards, an RNA adapter was ligated to the 5'-monophosphate of the RNA. First-strand cDNA synthesis was performed using an oligo(dT)-adapter primer and the M-MLV reverse transcriptase. The resulting cDNAs were PCR-amplified to about 10-20 ng/μL using a high-fidelity DNA polymerase. The cDNAs were purified using the Agencourt AMPure XP kit (Beckman Coulter Genomics) and were analysed by capillary electrophoresis. dRNA-Seq was undertaken to complement RNA-Seq data in the identification of transcript boundaries and operons (Sharma and Vogel 2014).

### 6.2.9 Pooled depletion of ribosomal RNA and cDNA size fractionation

Vertis Biotechnologie AG (Freising, Germany) employs an in-house ribosomal removal protocol for bacterial samples. This protocol has been optimised for ribosomal depletion

of individual samples. Given the high number of samples that were being sequenced here, we trialled a pooled ribosomal depletion strategy for these samples for cost saving purposes. From the whole transcriptome cDNA libraries, two approximately equimolar sub-pools were formed combining the samples by serotype. From the cDNA pools, cDNA molecules derived from ribosomal RNA structures were depleted using an in-house developed protocol. For Illumina NextSeq sequencing, the cDNA pools were size fractionated in the size range of 200 – 500 bp using preparative agarose gels. Aliquots of the size fractionated pools were analysed by capillary electrophoresis. The cDNAs from +/- TEX treated RNAs were added in approximately equimolar ratios prior to sequencing. The primers used for PCR amplification were designed for TruSeq sequencing according to the instructions of Illumina. The cDNA pools were single-read sequenced on an Illumina NextSeq 500 system using 75 bp read length.

#### 6.2.10 Transcripts mapping and counting

To enable inter-strain RNA-Seq analysis processing, the [multi-bacpipe pipeline](#) was applied to trimmed reads. Multi-bacpipe used STAR (v 2.7.6a) to align trimmed reads to reference genome sequences of BHN418UCL and BVJ1JL. Samtools (v1.11) (Li *et al.* 2009) was used to convert SAM files to BAM files and to sort and index BAM files for visualisation in IGV (Nicol *et al.* 2009). FeatureCount was applied to the resulting BAM files to evaluate strandedness of aligned transcripts and to generate transcripts count matrix (Liao, Smyth, and Shi 2014). Read counts were normalised as transcripts per million (TPM) to identify highly expressed and lowly expressed transcripts.

#### 6.2.11 Differential gene expression analysis

The generated transcripts count matrix was imported in R (v4.1.3) and normalised using DESeq2 Bioconductor package (v1.32.0). Normalisation of transcripts that mapped to BHN418UCL and BVJ1JL was performed separately. The group considerations for transcriptomic comparisons are illustrated in Figure (**Figure 6.2**). The log transformation within DESeq2 was applied to enable clustering and PCA analysis. PCA analysis was used to assess quality of replicates for each condition and evaluate batch effects. Unless otherwise stated, transcripts with fold change (FC) >1.5 and false discovery rate (FDR)

<0.05 relative to the control samples were considered as differentially expressed genes (DEGs).

## Transcriptomic Comparisons: Iron

2 x 2 x 2 x 2

Strains	Conditions	Time Points	Media
Serotype 1	Iron deplete	10 mins	CDM
Serotype 6B	Iron replete	50 mins	THY

**Figure 6.2: Illustration showing transcriptomic comparisons of iron experiment.**

Transcriptomic data was generated for serotype 1 (BVJ1JL) and serotype 6B (BHN418) grown in iron-deplete (induced by addition of 2,2 bipyridyl in media) and iron replete-conditions (induced by addition of  $\text{FeSO}_4$  in media). For both conditions, *S. pneumoniae* transcriptomic adaptation at 10 minutes and 50 minutes was evaluated. Experiments were performed in both enrich media Todd-Hewitt broth + Yeast extract (THY) and a chemically defined medium (CDM). THY media was considered an already iron-replete medium and therefore not supplemented with  $\text{FeSO}_4$ .

### 6.2.12 Statistics and visualization

All statistical analysis and visualizations were performed using latest Bioconductor packages in R environment (v4.1.3). T-test analysis of counts data was performed on DESeq2 normalised counts. Adjusted P-values <0.05 between group comparisons were considered significant.

### 6.2.13 Orthologous genes analysis

To enable comparison of transcriptomic results of BVJ1JL and BHN418UCL to other *S. pneumoniae* strains, genome sequences of six reference *S. pneumoniae* genome sequences downloaded from NCBI were used to identify orthologs of *S. pneumoniae* (Chapter 4). Prokka (v 1.14.6) was used to re-annotate genomes with NCBI GTF files and Roary (v 3.13.0) was used to identify orthologs based on these annotations. The reference genome sequences used for orthologs analysis and the rationale for choosing them are summarised in **Table 4.2** (Chapter 4).

### 6.2.14 Pathway analysis

Annotated pathways are available for a limited number of *S. pneumoniae* strains. To allow interrogation of biological pathways that may be enriched in study strains, the orthologs of KEGG annotated strains 670-6B (serotype 6B) and INV104 (serotype 1) were used for BHN418UCL (serotype 6B) and BVJ1JL (serotype 1) respectively. The Bioconductor R package clusterProfiler (Yu *et al.* 2012) was used for pathways analysis using BHN418UCL and BVJ1JL orthologs present in the KEGG annotated strains (Kanehisa 2000).

### 6.2.15 Creation of genome browser of reads

The multi-bacpipe pipeline generates bigWig file formats which enable visualisation of dense genomic data in a genome browser. The bigWig files generated were shared with Mr. Yan Li from the laboratory of Prof. Jay Hinton (University of Liverpool) to construct a prototype browser of transcriptomic reads. The genome browsers for visualisation of [Serotype 6B](#) and [Serotype 1](#) transcriptomic data have been made publicly accessible.

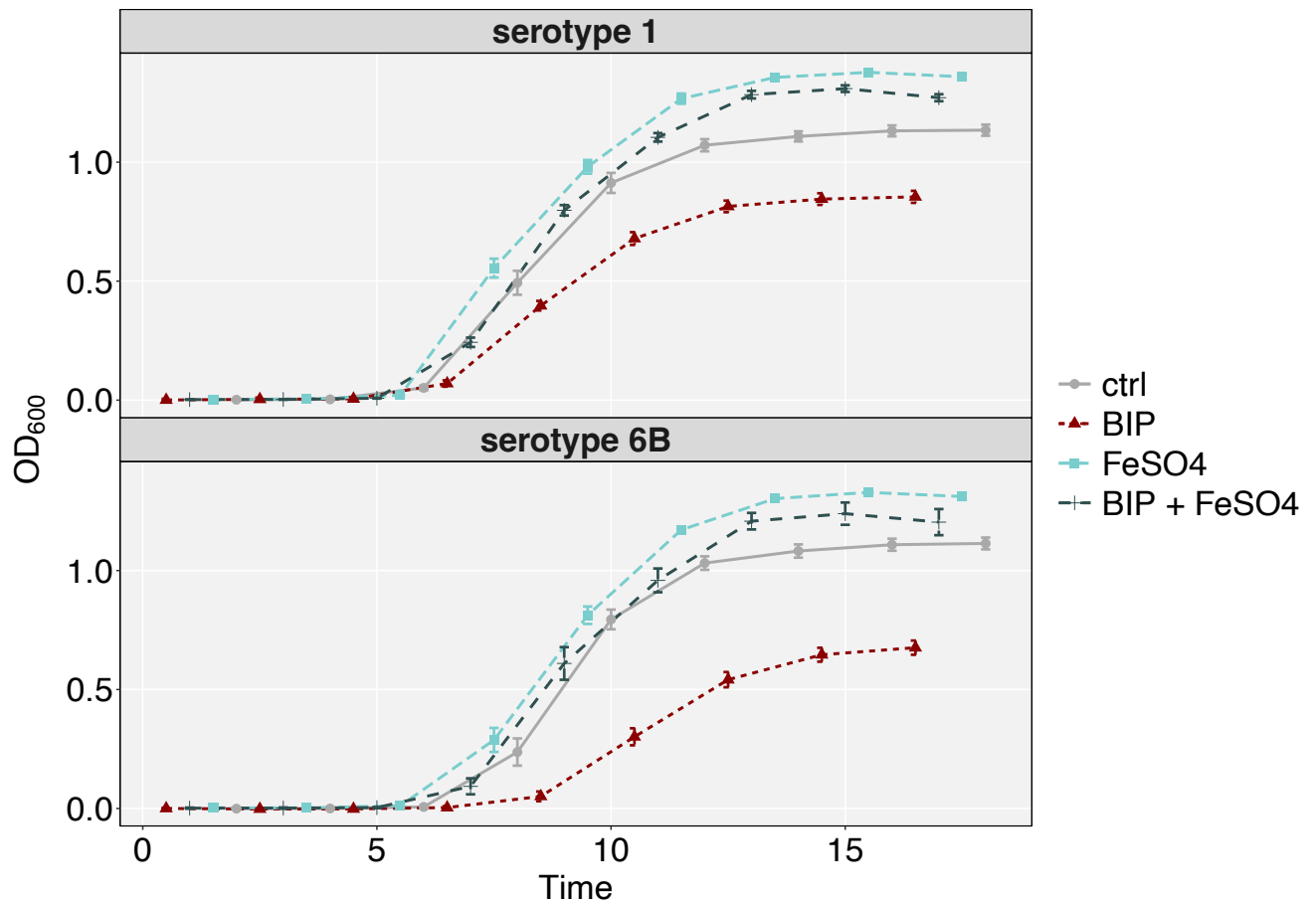
## 6.3 Results

### 6.3.1 Effects of iron-deplete and iron-replete conditions on *S. pneumoniae* growth

In chapter 3, it was shown that reduced availability of transition metals by chelation with 2,2 bipyridyl (BIP) reduced the growth of *S. pneumoniae* in a dose-dependent manner. Here, we investigated the effect of both transition metals depletion and repletion on *S. pneumoniae* serotype 1 (BVJ1JL) and serotype 6B (BHN418) growth in CDM. For both the serotype 1 (BVJ1JL) and serotype 6B (BHN418) strains, transition metals stress resulted in growth defect demonstrated by longer lag phase, lower growth density, and lower maximal growth (**Figure 6.3A**). Iron repletion (by addition of  $\text{FeSO}_4$ ) accelerated growth in both strains and was able to rescue growth defects caused by addition of BIP (**Figure 6.3A**). This data demonstrates that transition metal availability has considerable effect on *S. pneumoniae* proliferation which may contribute to pathogenesis. To investigate whether *S. pneumoniae* growth rescue was iron specific, we substitute iron with manganese in presence of BIP. Addition of manganese into growth cultures also rescued bacterial growth in a similar manner to iron suggesting that *S. pneumoniae* may be able to substitute manganese and iron for similar biological processes (**Figure 6.3B**). These results corroborate the findings of Cao *et al.*, (Cao *et al.* 2018), which found that supplementing Mn-deficient medium with Fe recovers the growth of *S. pneumoniae*.

## FeSO<sub>4</sub> Rescues Growth Defects Caused by 2,2'-Bipyridyl

2,2'-Bipyridyl final concentrations 0.5mM; FeSO<sub>4</sub> final concentrations 0.3mM

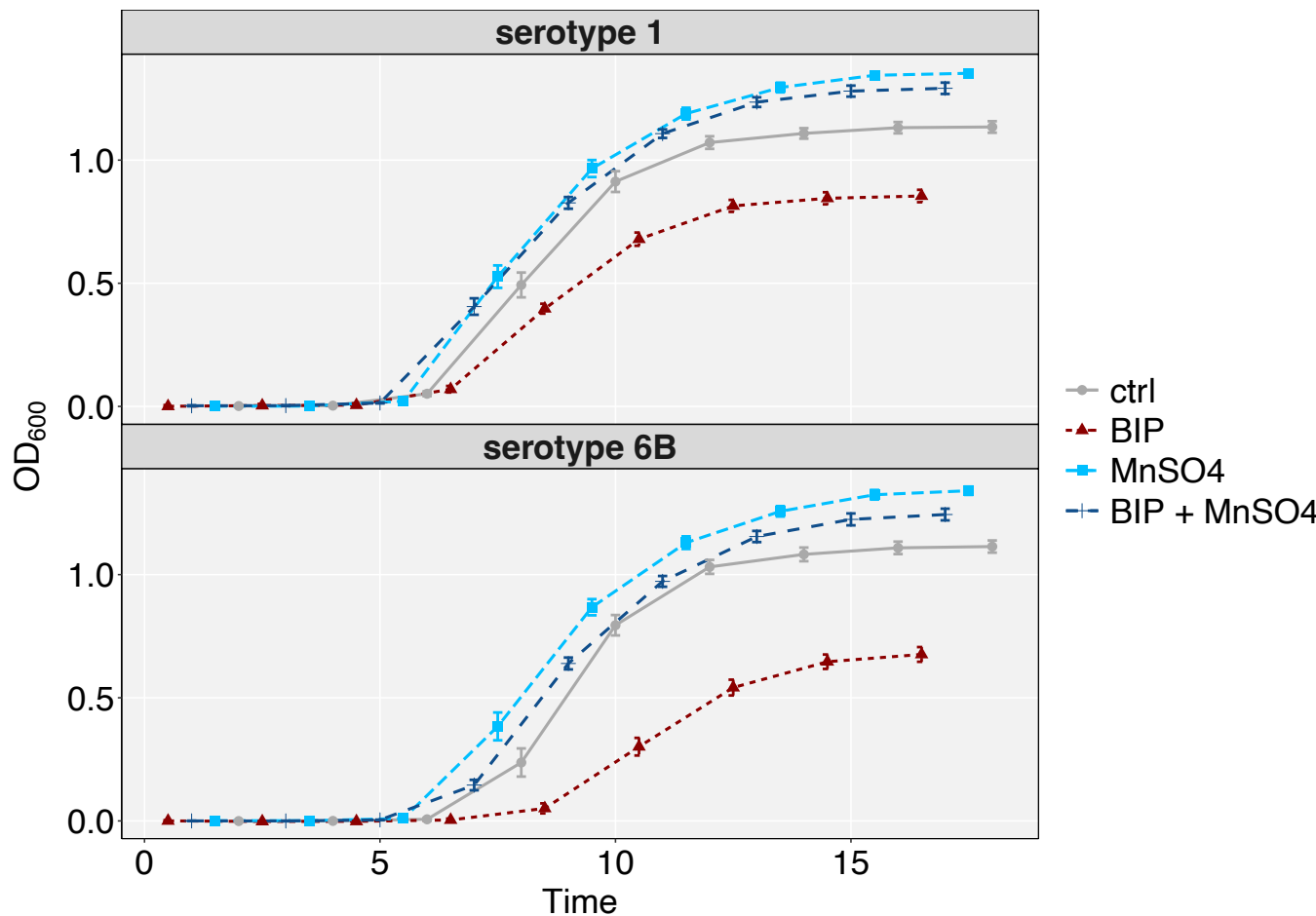


**Figure 6.3A: Chelation of transition metal ions impaired growth of *S. pneumoniae* strains BVJ1JL (Serotype 1) and BHN418 (serotype 6B) in CDM.** Chelation with 2,2' bipyridyl (BIP) extended *S. pneumoniae* lag growth phase and reduced maximal growth of serotype 1 and serotype 6B. Iron repletion (FeSO<sub>4</sub>) had an opposite effect and accelerated *S. pneumoniae* growth. *S. pneumoniae* growth defects caused by chelation were also rescued by iron repletion (BIP + FeSO<sub>4</sub>) to controls levels (ctrl). *S. pneumoniae* growth in CDM was measured by optical density over an 18hrs incubation period using a plate reader.



## MnSO<sub>4</sub> Rescues Growth Defects Caused by 2,2'-Bipyridyl

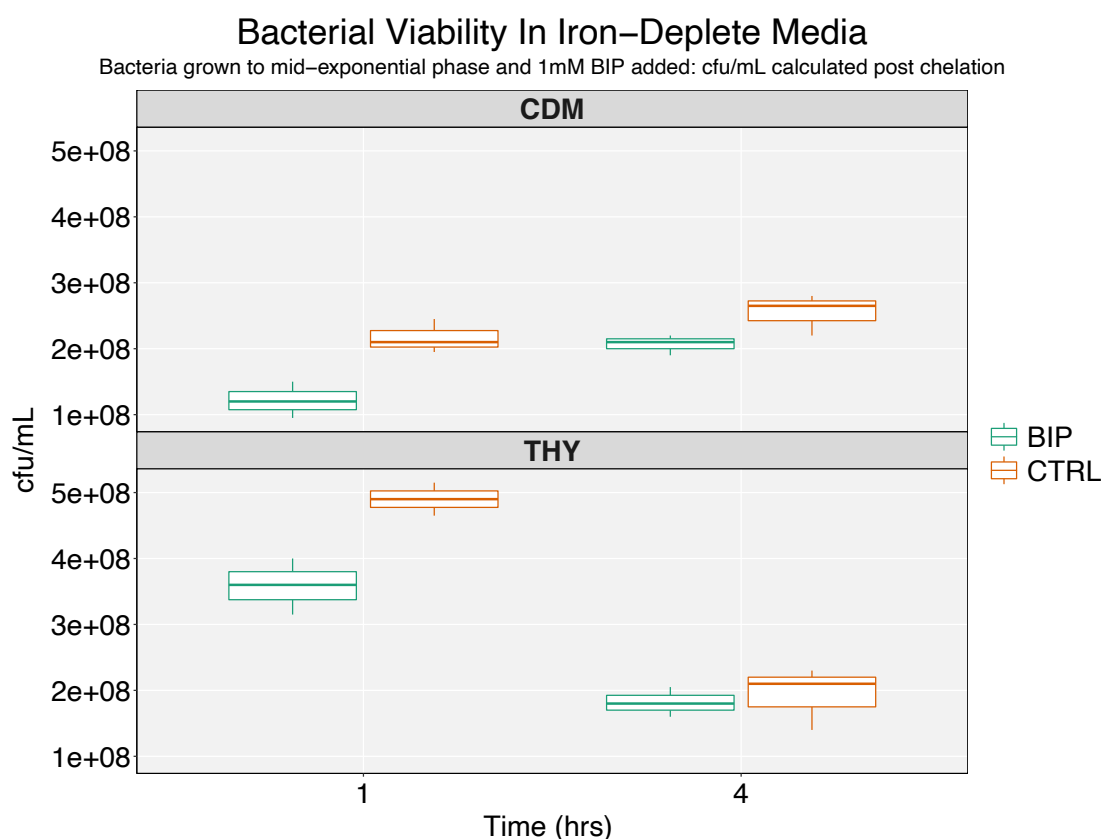
2,2'-Bipyridyl final concentrations 0.5mM; MnSO<sub>4</sub> final concentrations 0.3mM



**Figure 6.3B: Manganese repletion accelerated growth of *S. pneumoniae* strains BVJ1JL (Serotype 1) and BHN418UCL (serotype 6B) similar to iron repletion.** Chelation with 2,2' bipyridyl extended *S. pneumoniae* lag growth phase and reduced maximal growth of serotype 1 (BVJ1JL) and serotype 6B (BIP). Manganese repletion (MnSO<sub>4</sub>) had an opposite effect and accelerated *S. pneumoniae* growth. *S. pneumoniae* growth defects caused by chelation were also rescued by Manganese repletion (BIP + MnSO<sub>4</sub>) to controls levels (ctrl). *S. pneumoniae* growth in CDM was measured by optical density over an 18hrs incubation period using a plate reader.

### 6.3.2 Assessing potential toxic effects of 2,2' bipyridyl on *S. pneumoniae* viability

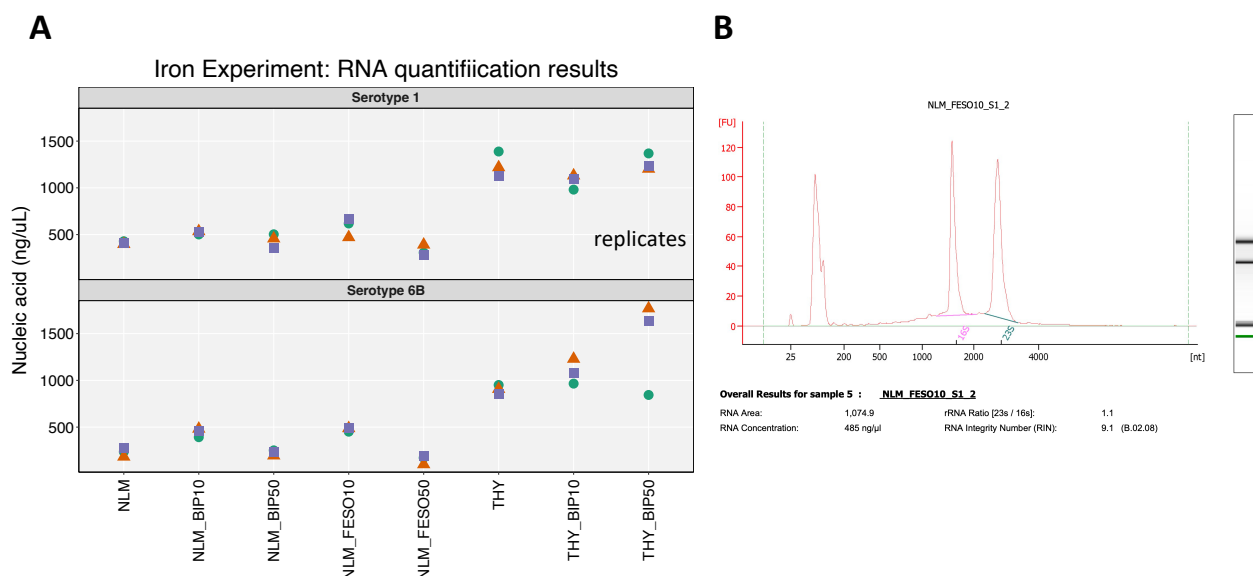
Similar to the results of section 6.3.1, the addition of 1mM BIP to MEP *S. pneumoniae* (serotype 1) cultures reduced bacterial growth as demonstrated by lower CFU/mL after 1hr and 4hrs in both THY and CDM (**Figure 6.4**). This highlights the importance of metal ions availability for normal *S. pneumoniae* growth. However, the difference in CFU/mL between growth cultures in CDM with and without BIP was within the same log ( $10^{-8}$  cells) (**Figure 6.4**). These results were taken to suggest that 1mM BIP is non-toxic to *S. pneumoniae* viability, and the bacteria continued to adapt and proliferate in this environment for at least 4hrs albeit at a slower rate. *S. pneumoniae* growth in THY showed reduced CFU/mL in both test and control samples after the 4hrs timepoint (approximately 8 hours from start of cultures), which was expected due to autolysis (**Figure 6.4**). *S. pneumoniae* reaches stationary phase at approximately 7hrs during growth in THY (Chapter 3). To further assess potential effects of BIP on *S. pneumoniae* growth, the cultures were plated on CBA plate and incubated overnight in standard conditions. Visual inspections of bacterial growth in CBA showed slightly smaller colonies in BIP supplemented cultures at both the 1hr and 4hr timepoints, in line with a slower growth rate, but no other morphology differences were observed compared to controls. Therefore, the 1mM concentration of BIP used was considered to be non-toxic and transcriptomics experiments were designed to investigate the gene expression changes during adaptation in these conditions.



**Figure 6.4:** *S. pneumoniae* serotype 1 (BVJ1JL) remained viable in 1mM concentrations of 2,2' bipyridyl (BIP) up to 4 hours in CDM and THY. 1mM BIP was added to BVJ1JL (serotype 1) culture grown to MEP in Todd-Hewitt broth + Yeast (THY) and a chemically defined media (CDM). The number of colony forming units (CFU/mL) were measured at 1hr and 4hrs post addition of BIP and compared to growth in controls not supplemented with BIP.

### 6.3.3 High yield RNA quantities generated

The optimised RNA isolation method used (Chapter 3) resulted in high bacterial RNA quantities ranging from 101.8ng/μL to 1769.2ng/μL (**Figure 6.5A**). Examination of RNA quality by capillary electrophoresis (Agilent Bioanalyzer2100) indicated intact prokaryotic RNA with an average RNA Integrity Number (RIN) score of 9 (**Figure 6.5B**). Experiments conducted in THY had an average of > two-fold higher RNA quantities compared to experiments in CDM (**Figure 6.5A**).



**Figure 6.5. High yield RNA quantities generated for iron experiment. A.** RNA quantification results of three replicates submitted for sequencing. RNA quantification data was generated with Nanodrop and validated using Qubit. The colour and shapes correspond to the replicate numbers **B.** A representative capillary electrophoresis data of an experimental replicate showing distinct bacterial 16S and 23S ribosomal RNAs bands indicative of intact RNA.

### 6.3.4 Total transcriptomics reads

The total number of raw reads generated for all samples in this experiment was 988,176,266 reads resulting in an average of ~20 million reads per library (range 12.2M to 21.8M). Reads that passed QC were subsequently mapped to the respective *S. pneumoniae* reference genomes. The pooled sample ribodepletion strategy that was attempted to decrease cost did not yield optimal removal of bacterial ribosomal RNA, with as many as 46% of reads in some samples being ribosomal sequences (**Table 6.4**). In comparison, individual rRNA depletion of bacterial samples can achieve near complete removal (<0.15%) of rRNA encoding genes (Haas *et al.* 2012). Ribosomal RNA quantities were relatively less abundant in samples of BVJ1JL (serotype 1) compared to BHN418UCL (serotype 6B) (**Table 6.4**), suggesting potential avenues for further optimisation of the protocol to enable more efficient ribodepletion in the future. However, despite the relatively high number of ribosomal reads of some libraries, an average of 10 million (M) reads were assigned as mRNAs fragments across all libraries (range 5.1M -15M) (**Table**

**6.4).** For bacterial RNA-Seq data with well correlated replicates, non-rRNA fragments of 2-3M have been determined to be adequate for differential gene expression analysis (Haas *et al.* 2012). Ribodepletion of dRNA-Seq libraries is not a requirement and the low percentage of assigned mRNA reads in these libraries (libraries with prefix “Pool” in **Table 6.4**) were as expected (Sharma and Vogel 2014).

**Table 6-4:** Mapping results of RNA-Seq data to *S. pneumoniae* reference genome sequences BHN418UCL (serotype 6B) and BVJ1JL (serotype 1) showing the amount of rRNA sequences, number of mapped, and unmapped mRNA reads in each library.

Sample	Reference strain	All (%)	rRNA (%)	Unmapped (%)	Multimap (%)	Assigned (%)
CDM_6B_1	BHN418UCL	15329221 (100%)	6344366 (41.38)	379799 (2.47)	98140 (0.64)	7779625 (50.75)
CDM_6B_2	BHN418UCL	15616753 (100%)	5774218 (36.97)	421492 (2.70)	110242 (0.71)	8543033 (54.70)
CDM_6B_3	BHN418UCL	15870169 (100%)	6144033 (38.71)	409868 (2.58)	103416 (0.65)	8401670 (52.94)
CDM_BIP10_6B_1	BHN418UCL	16886511 (100%)	5382453 (31.87)	451045 (2.67)	104965 (0.62)	9830252 (58.21)
CDM_BIP10_6B_2	BHN418UCL	19627636 (100%)	7715171 (39.31)	448986 (2.29)	104444 (0.53)	10150092 (51.71)
CDM_BIP10_6B_3	BHN418UCL	17432416 (100%)	6692405 (38.40)	398243 (2.28)	93221 (0.54)	9219567 (52.89)
CDM_BIP10_S1_1	BVJ1JL	18866058 (100%)	4052707 (21.48)	75216 (0.40)	67657 (0.36)	13878108 (73.56)
CDM_BIP10_S1_2	BVJ1JL	15751511 (100%)	4084444 (25.93)	87138 (0.55)	54641 (0.35)	10742390 (68.20)
CDM_BIP10_S1_3	BVJ1JL	16334163 (100%)	3753022 (22.98)	87006 (0.53)	57310 (0.35)	11695802 (71.60)
CDM_BIP50_6B_1	BHN418UCL	16150276 (100%)	6313366 (39.09)	406962 (2.52)	105812 (0.66)	8491962 (52.58)
CDM_BIP50_6B_2	BHN418UCL	16666165 (100%)	6634016 (39.80)	420251 (2.52)	111481 (0.67)	8708676 (52.25)

(Table continued)

<b>CDM_BIP50_6B_3</b>	BHN418UCL	15934679 (100%)	6216146 (39.01)	399250 (2.51)	99254 (0.62)	8392903 (52.67)
<b>CDM_BIP50_S1_1</b>	BVJ1JL	21049453 (100%)	4937999 (23.46)	105565 (0.50)	77915 (0.37)	15097414 (71.72)
<b>CDM_BIP50_S1_2</b>	BVJ1JL	17378146 (100%)	4580631 (26.36)	91493 (0.53)	61524 (0.35)	11923002 (68.61)
<b>CDM_BIP50_S1_3</b>	BVJ1JL	14129996 (100%)	4600375 (32.56)	74945 (0.53)	47853 (0.34)	8856299 (62.68)
<b>CDM_FESO10_6B_1</b>	BHN418UCL	16187819 (100%)	6053631 (37.40)	429324 (2.65)	102074 (0.63)	8778665 (54.23)
<b>CDM_FESO10_6B_2</b>	BHN418UCL	18413184 (100%)	6962149 (37.81)	467025 (2.54)	114506 (0.62)	9855816 (53.53)
<b>CDM_FESO10_6B_3</b>	BHN418UCL	14338152 (100%)	5233529 (36.50)	395629 (2.76)	93400 (0.65)	7919214 (55.23)
<b>CDM_FESO10_S1_1</b>	BVJ1JL	13865703 (100%)	3225663 (23.26)	71293 (0.51)	49964 (0.36)	9906155 (71.44)
<b>CDM_FESO10_S1_2</b>	BVJ1JL	15275443 (100%)	3364743 (22.03)	72906 (0.48)	55776 (0.37)	11089890 (72.60)
<b>CDM_FESO10_S1_3</b>	BVJ1JL	16167684 (100%)	4723716 (29.22)	86281 (0.53)	54865 (0.34)	10567932 (65.37)
<b>CDM_FESO50_6B_1</b>	BHN418UCL	18625400 (100%)	7053176 (37.87)	489928 (2.63)	129404 (0.70)	10169400 (54.60)
<b>CDM_FESO50_6B_2</b>	BHN418UCL	17346821 (100%)	6227756 (35.90)	470844 (2.71)	125848 (0.73)	9767689 (56.31)
<b>CDM_FESO50_6B_3</b>	BHN418UCL	16743853 (100%)	6539176 (39.05)	436118 (2.61)	104036 (0.62)	9079254 (54.22)

*(Table continued)*

<b>CDM</b> <b>_FESO50_S1_1</b>	BVJ1JL	16458055 (100%)	3793529 (23.05)	91523 (0.56)	66550 (0.40)	11840324 (71.94)
<b>CDM</b> <b>_FESO50_S1_2</b>	BVJ1JL	16306224 (100%)	3442174 (21.11)	73173 (0.45)	65686 (0.40)	12021597 (73.72)
<b>CDM</b> <b>_FESO50_S1_3</b>	BVJ1JL	18946343 (100%)	3949628 (20.85)	95114 (0.50)	76331 (0.40)	14038747 (74.10)
<b>CDM_S1_1</b>	BVJ1JL	16822377 (100%)	4695864 (27.91)	77427 (0.46)	60896 (0.36)	11348054 (67.46)
<b>CDM_S1_2</b>	BVJ1JL	14746753 (100%)	3952502 (26.80)	65265 (0.44)	51806 (0.35)	10110438 (68.56)
<b>CDM_S1_3</b>	BVJ1JL	16397378 (100%)	4188712 (25.55)	65012 (0.40)	57903 (0.35)	11470018 (69.95)
<b>PoolSerotyp1</b>	BVJ1JL	17323709 (100%)	14462204 (83.48)	1164075 (6 .72)	9048 (0.05)	1613341 (9.31)
<b>PoolSerotyp1_TEX_</b> <b>S1</b>	BVJ1JL	17945850 (100%)	12556998 (69.97)	3053079 (17.01)	18287 (0.10)	1987283 (11.07)
<b>PoolSerotyp6B</b>	BHN418UCL	14291682 (100%)	11170691 (78.16)	1297404 (9.08)	16129 (0.11)	1712266 (11.98)
<b>PoolSerotyp6B_TEX</b>	BHN418UCL	19744338 (100%)	13281921 (67.27)	3469784 (17.57)	29766 (0.15)	2557977 (12.96)
<b>THY_6B_1</b>	BHN418UCL	12503590 (100%)	5586781 (44.68)	1208095 (9.66)	27110 (0.22)	5240995 (41.916)
<b>THY_6B_2</b>	BHN418UCL	11865421 (100%)	5142165 (43.34)	1112625 (9.38)	26922 (0.23)	5138718 (43.31)
<b>THY_6B_3</b>	BHN418UCL	14082159 (100%)	6336880 (44.99)	1250229 (8.88)	31142 (0.22)	5974559 (42.43)
<i>(Table continued)</i>						



<b>THY_BIP10_6B_1</b>	BHN418UCL	13047222 (100%)	6052408 (46.39)	1047481 (8.03)	28635 (0.22)	5465210 (41.89)
<b>THY_BIP10_6B_2</b>	BHN418UCL	13361105 (100%)	6197208 (46.38)	1208038 (9.04)	28238 (0.21)	5468640 (40.93)
<b>THY_BIP10_6B_3</b>	BHN418UCL	15026520 (100%)	6784741 (45.15)	1186661 (7.90)	32900 (0.22)	6498438 (43.24)
<b>THY_BIP10_S1_1</b>	BVJ1JL	15735236 (100%)	4907502 (31.19)	1071037 (6.80)	32569 (0.20)	9182185 (58.35)
<b>THY_BIP10_S1_2</b>	BVJ1JL	18097160 (100%)	5094443 (28.15)	1042478 (5.76)	41205 (0.23)	11263868 (62.24)
<b>THY_BIP10_S1_3</b>	BVJ1JL	19911233 (100%)	7166945 (35.99)	1341808 (6.74)	39596 (0.20)	10741107 (53.95)
<b>THY_BIP50_6B_1</b>	BHN418UCL	13666685 (100%)	7047337 (51.57)	771673 (5.65)	27904 (0.20)	5343964 (39.10)
<b>THY_BIP50_6B_2</b>	BHN418UCL	16699225 (100%)	7570506 (45.33)	673885 (4.04)	37590 (0.23)	7698686 (46.10)
<b>THY_BIP50_6B_3</b>	BHN418UCL	13685391 (100%)	5925583 (43.29)	523557 (3.83)	31700 (0.23)	6604629 (48.26)
<b>THY_BIP50_S1_1</b>	BVJ1JL	18402977 (100%)	6403838 (34.79)	804327 (4.37)	39548 (0.22)	10493182 (57.02)
<b>THY_BIP50_S1_2</b>	BVJ1JL	13143765 (100%)	4991128 (37.97)	720318 (5.48)	26021 (0.20)	6951736 (52.89)
<b>THY_BIP50_S1_3</b>	BVJ1JL	12113529 (100%)	4712887 (38.90)	564699 (4.66)	23119 (0.19)	6385849 (52.71)
<b>THY_S1_1</b>	BVJ1JL	17000523 (100%)	5697664 (33.52)	1829060 (10.76)	31000 (0.18)	8872894 (52.19)

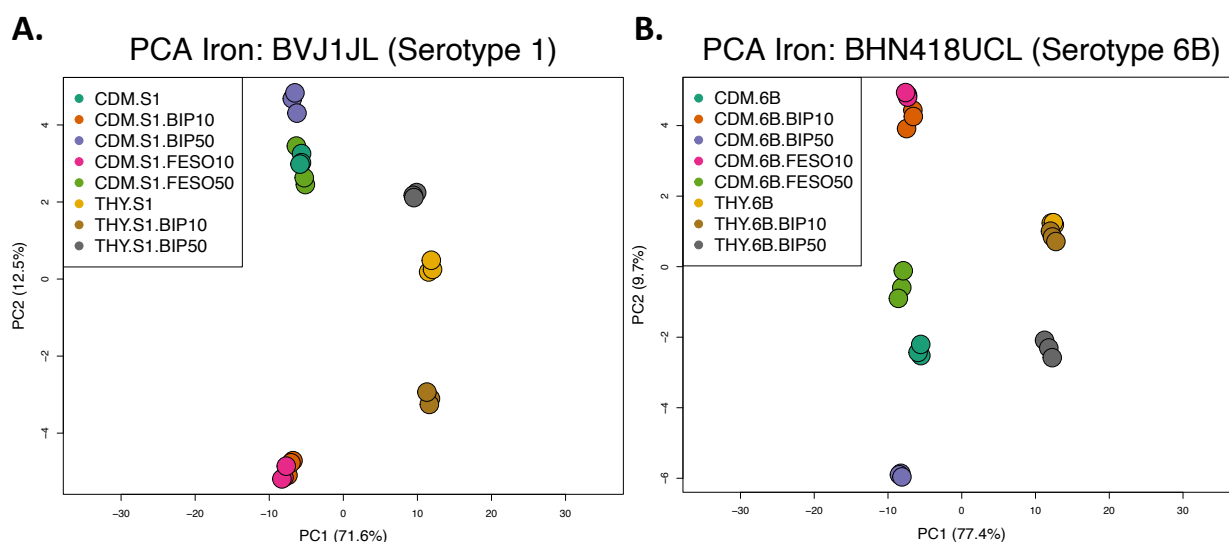
*(Table continued)*

THY_S1_2	BVJ1JL	16285580	5204651	1257807	34078	9238648
		(100%)	(31.96)	(7.72)	(0.21)	(56.73)
THY_S1_3	BVJ1JL	14379791	4392268	1099921	30902	8341627
		(100%)	(30.55)	(7.65)	(0.22)	(58.01)

### 6.3.5 RNA-Seq data quality

Principal component analysis (PCA) of log transformed count data in DESeq2 indicated high quality transcriptomic data with highly correlated biological replicates clustering by experimental condition (**Figure 6.6**). Evidence of experimental batch effects was absent from the PCA analysis. For both BHN418UCL (serotype 6B) and BVJ1JL (serotype 1) transcriptomics datasets, the longest distance between points along PC1 (PC with highest variance) was between experiments conducted in THY and CDM (**Figure 6-6**).

## PCA Biplots of BVJ1JL and BHN418UCL in Iron-deplete and Iron-replete Conditions



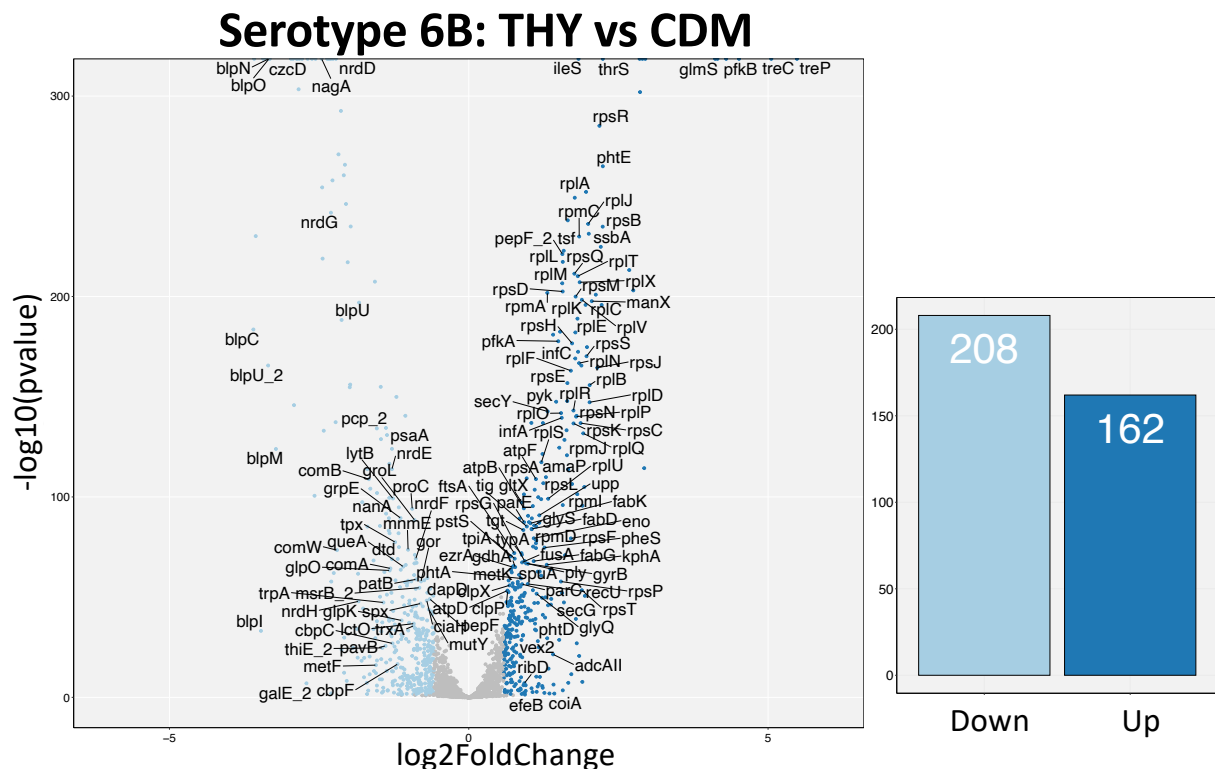
**Figure 6.6: Principal component analysis (PCA) of RNA-seq data showed high reproducibility of biological replicates.** Replicates of each experimental condition are coloured uniquely **A.** PCA clustering of conditions and biological replicates for BVJ1JL (serotype 1) **B.** PCA clustering of replicates for BHN418UCL (serotype 6B) in THY and CDM.

### 6.3.6 Analysis of *S. pneumoniae* orthologs

Using Roary, the pangenome of the six reference genome sequences together with BHN418UCL (serotype 6B) and BVJ1JL (serotype 1) was determined to consist of 3033 genes in total. Of these 1473 genes were core and present in all eight genome sequences. Analysis of BHN418UCL and BVJ1JL genome assemblies resulted in a pangenome of 2338 genes with core genome between the strains consisting of 1674 genes. This is consistent with what has been previously described (Donati *et al.* 2010; Kulohoma *et al.* 2015).

### 6.3.7 Impact of growth media on *S. pneumoniae* transcriptome

Gene expression of MEP control samples grown in THY and CDM without supplementation with BIP or FeSO<sub>4</sub> were compared for serotype 1 (BVJ1JL) and serotype 6B (BHN418). These analyses indicated substantial differences in *S. pneumoniae* gene expression with 370 and 363 genes differentially regulated by BHN418 (serotype 6B) and BVJ1JL (serotype 1) respectively (using FC >2.0 and FDR <0.05) (**Figure 6.7A** and **Figure 6.7B**). Using CDM transcriptome at MEP as a reference level, genes downregulated in THY for both strains included pneumococcal virulence factors (*psaA*, *nanA*, *pavB*, and *lytA*), two-component system genes (e.g., *comA*, *comW*), bacteriocin like-peptides (e.g., *blpM*, *blpN*, *blpO*, *blpU*, *blpC*) as well as the zinc efflux system gene *czcD*. Upregulated genes in THY included several ribosomal genes (e.g., *rpsM*, *rpsB*, *rplA*, *rpsR*), pneumolysin (*ply*), carbohydrate metabolism genes (e.g., *manX*, *lacA*, *lacB*), and metal ion response genes *adcAII* and *adcR*. These analyses highlighted the nutritional differences between THY and CDM and the influence of growth media on *S. pneumoniae* transcriptome (**Figure 6.7A** and **Figure 6.7B**). Although experiments were sampled at MEP of growth in both CDM and THY, the growth trajectory in the two media types is different (Chapter 3) and it is plausible that some of the transcriptomics differences observed may be growth related. Considering the confounding effects of growth medium, inter- and intrastrain transcriptomic comparisons were confined to experiments conducted with the same growth media type.



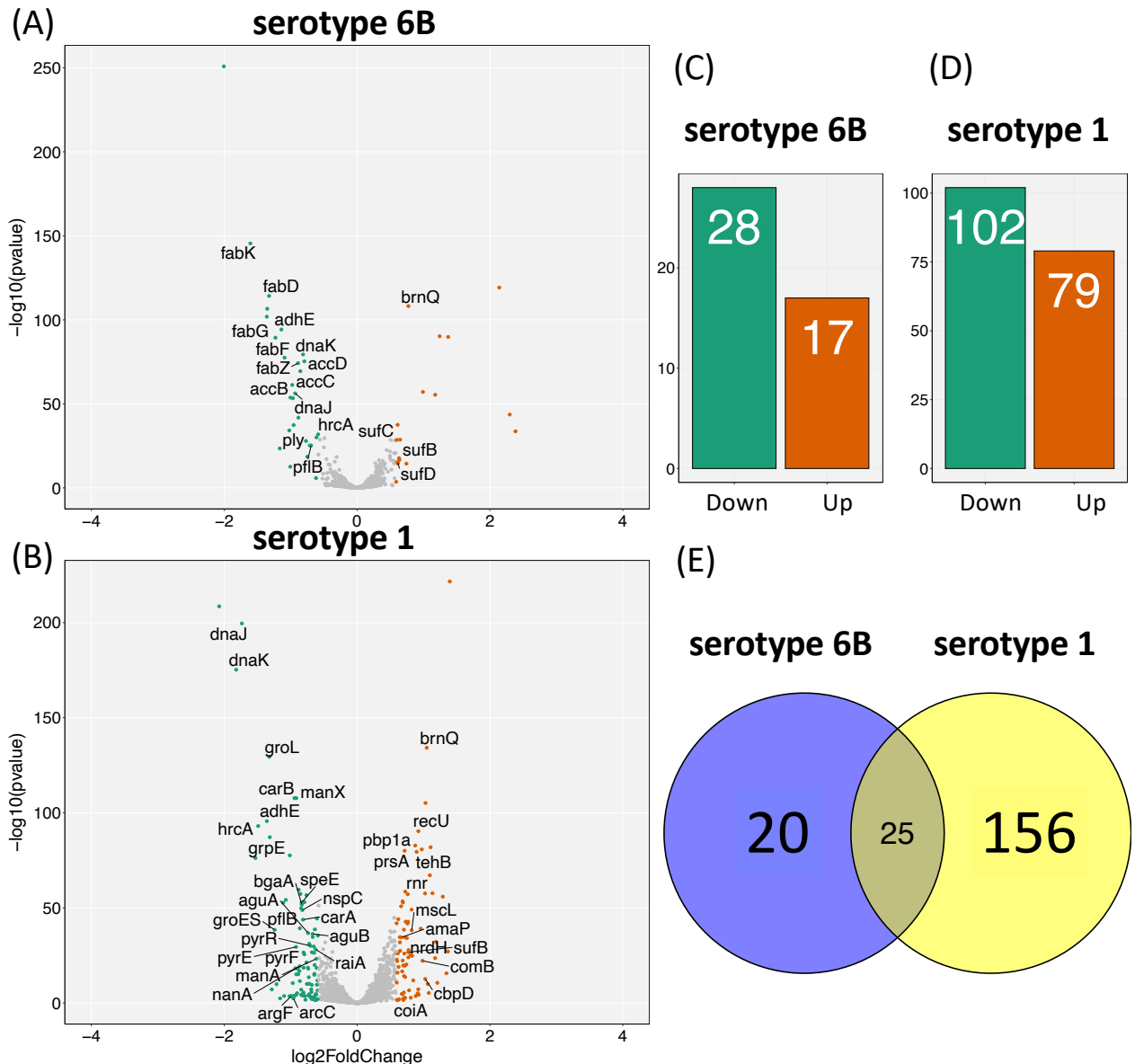
**Figure 6.7A: Differentially regulated genes of *S. pneumoniae* strain BHN418 (serotype 6B) grown in Todd Hewitt broth + yeast extract (THY) and a chemically defined media (CDM).** *S. pneumoniae* was grown to mid exponential phase (MEP) in THY and CDM base media without supplementation. Volcano and bar plots show the distribution of up- (dark blue colour) and down-regulated (light blue colour) genes. The gene expression in CDM was used as reference level for transcriptomic analyses. Points in dark blue are upregulated genes in THY compared to CDM and points in light blue are genes that are upregulated in CDM compared to THY. Genes with fold-changes  $>2$  and  $\text{FDR} < 0.05$  were classified as DEGs in this analysis.



### 6.3.8 Early transcriptomic response of *S. pneumoniae* during iron limitation in THY

Induction of iron limitation by addition of 1mM BIP in THY resulted in more downregulated genes than upregulated genes for both serotype 1 (BVJ1JL) and serotype 6B (BHN418) strains after 10 minutes (**Figure 6.8**). Of the 201 differentially expressed genes (DEGs) between the strains, 12.4% (25/201) were shared, suggesting mainly strain-specific differences in gene regulation in this environment (**Figure 6.8E**). All the genes upregulated by serotype 6B (n =17) specifically encode transition metal response genes (**Table 6.5A**). These included the complete operon of the *S. pneumoniae* iron uptake (*piu*) genes *piuB*, *piuC*, *piuD*, and *piuA*, and genes encoding pneumococcal iron-sulfur biogenesis system (*sufC*, *sufD*, *sufB*, *sufS*, *sufE*). Genes that form the specialised copper export system (*cop* operon) (Shafeeq *et al.* 2011) *copY*, *cupA*, and *copA* were also upregulated by serotype 6B at this timepoint. In comparison, the serotype 1 (BVJ1JL) strain upregulated and downregulated more than three times the number of DEG by BHN418, which included a more diverse set of genes and other ABC transporters of yet unknown function. The two most significantly upregulated genes by the serotype 1 strain were *cshA*, a DEAD-box ATP-dependent RNA helicase, and *brnQ*, a branched-chain amino acid transport carrier protein (**Table 6.5B**). The *piu* operon (*piuB*, *piuC*, *piuD*, and *piuA*) and the *cop* operon (*copY*, *cupA*, and *copA*) genes were also upregulated by the serotype 1 strain at this timepoint, similar to serotype 6B. The *gls24* gene (SPD\_1590 in D39 strain), which was previously annotated as a general response gene but recently suggested as a pneumococcal iron transporter in D39 strain (serotype 2) (Miao *et al.* 2018), was significantly upregulated by both our serotype 1 and serotype 6B strains, supporting sensitivity in the regulation of the gene to low iron levels (**Table 6.5A** and **Table 6.5B**). The serotype 6B strain downregulated expression of fatty acid synthesis genes operon (*fabZGDKF*, *accCB*) and sugar (PTS) transporters while downregulated genes in the serotype 1 strain were dominated by stress response chaperone proteins (e.g., *hrcA*, *grpE*, *groELS*, *dnaJ*, *dnaK*), several sugar (PTS) transporters (e.g., *bguD*, *JN057\_RS03195*, *manX*) and other diverse gene sets including virulence factor *nanA* (**Table 6.5C** and **Table 6.5D**). An iron-binding alcohol dehydrogenase, *adhE*, which is associated with *S. pneumoniae* adaptation to oxidative stress was significantly downregulated by both serotype 1 and serotype 6B strains (Andrews, Robinson, and Rodríguez-Quñones 2003; Echave *et al.* 2003).

## DGE at 10 mins (THY): Iron-deplete Conditions



**Figure 6.8: Differential gene expression (DGE) of BHN418 (serotype 6B) and BVJ1JL (serotype 1) after 10 minutes in iron-deplete environment in THY.** Volcano (A - B) and bar (C - D) plots showing up (brown) and down (green) regulated genes in serotype 6B and serotype 1 strains. Each point in the volcano plots indicate a DEG with colour indicative of expression direction (brown = upregulated, green = downregulated). Where available in the original annotation file, short gene names are annotated on volcano plots. (E) Venn diagram showing overlap in the number of DEGs between serotype 6B and serotype 1 at this environment. DEGs were defined by fold change  $>1.5$  and false discovery rate  $\text{FDR} < 0.05$ .

**Table 6.5A:** Seventeen differentially upregulated genes by *S. pneumoniae* serotype 6B (BHN418) after 10 minutes growth in iron-deplete THY (1mM BIP). Genes were ordered by statistical significance (P-adj). Top upregulated genes in both strains are in bold. RefSeq locus tags of the published sequence of BVJ1JL are shown (Betts *et al.* 2021).

locus	Gene	FC	P-adj	Annotation
<b>Serotype 6B (BHN418)</b>				
JN057_RS03550	<i>copA</i>	4.41	5.81E-120	Heavy metal translocating P-type ATPase
JN057_RS03085	<i>brnQ</i>	1.71	6.39E-109	branched-chain amino acid transport system II carrier protein
JN057_RS09010	<i>piuC</i>	2.37	5.81E-91	Iron compound ABC uptake transporter permease protein PiuC
JN057_RS09005	<i>piuB</i>	2.59	1.41E-90	Iron compound ABC uptake transporter permease protein PiuB
JN057_RS09020	<i>piuA</i>	1.99	7.80E-58	Iron compound ABC uptake transporter substrate-binding protein PiuA
JN057_RS09015	<i>piuD</i>	2.26	4.08E-56	Iron compound ABC uptake transporter ATP-binding protein PiuD
JN057_RS03540	<i>copY</i>	4.92	2.36E-44	CopY/TcrY family copper transport repressor
<b>JN057_RS0866</b>	<b><i>gls24</i></b>	1.53	2.73E-38	Asp23/Gls24 family envelope stress response protein
<b>5</b>				
JN057_RS03545	<i>cupA</i>	5.22	2.19E-34	Copper-exporting ATPase
JN057_RS08670		1.57	2.08E-29	CsbD family protein
JN057_RS04165	<i>sufC</i>	1.51	2.63E-29	Fe-S cluster assembly ATPase SufC
JN057_RS07420		1.55	2.01E-18	CsbD family protein
JN057_RS04185	<i>sufB</i>	1.56	2.20E-17	Fe-S cluster assembly protein SufB
JN057_RS04170	<i>sufD</i>	1.53	5.59E-16	Fe-S cluster assembly protein SufD
JN057_RS04175	<i>sufS</i>	1.53	3.41E-15	cysteine desulfurase
JN057_RS04180	<i>sufE2</i>	1.67	3.73E-15	SUF system NifU family Fe-S cluster assembly protein
JN057_RS09745		1.50	2.42E-04	ABC transporter ATP-binding protein



**Table 6.5B:** Top 35 differentially upregulated genes by *S. pneumoniae* serotype 1 (BVJ1JL) after 10 minutes growth in iron-deplete THY (1mM BIP). Genes were ordered by statistical significance (P-adj). Top upregulated genes in both strains are in bold. RefSeq locus tags of the published sequence of BVJ1JL are shown (Betts *et al.* 2021).

locus	Gene	FC	P-adj	Annotation
<b>Serotype 1 (BVJ1JL)</b>				
<i>JN057_RS07680</i>	<i>cshA</i>	2.63	2.41E-222	DEAD/DEAH box helicase
<i>JN057_RS03085</i>	<i>brnQ</i>	2.07	6.63E-135	branched-chain amino acid transport system II carrier protein
<i>JN057_RS01190</i>		2.04	7.18E-106	SP_0198 family lipoprotein
<i>JN057_RS01995</i>	<i>recU</i>	1.90	4.67E-91	Holliday junction resolvase RecU
<i>JN057_RS01990</i>	<i>pbp1a</i>	1.84	1.57E-83	Penicillin-binding protein PBP1A
<i>JN057_RS07115</i>		2.15	1.26E-82	DUF1836 domain-containing protein
<i>JN057_RS04930</i>		1.965	1.76E-81	DUF1002 domain-containing protein
<i>JN057_RS04705</i>	<i>prsA</i>	1.64	8.75E-81	peptidylprolyl isomerase PrsA
<i>JN057_RS04685</i>	<i>tehB</i>	1.862	3.19E-80	SAM-dependent methyltransferase TehB
<i>JN057_RS07120</i>		2.14	6.80E-68	hemolysin III family protein
<i>JN057_RS03040</i>		1.65	2.92E-59	M57 family metalloprotease
<i>JN057_RS08275</i>	<i>gntR</i>	2.20	1.89E-58	GntR family transcriptional regulator
<i>JN057_RS08285</i>		2.032	2.43E-58	ABC transporter ATP-binding protein
<i>JN057_RS04675</i>	<i>rnr</i>	1.70	6.86E-58	ribonuclease R
<i>JN057_RS09685</i>	<i>diiA</i>	2.45	1.17E-56	DUF1542 domain-containing protein/ cell wall
<i>JN057_RS08495</i>		1.61	3.17E-54	aquaporin
<i>JN057_RS11050</i>		1.61	1.56E-53	ParB/RepB/Spo0J family partition protein
<i>JN057_RS01185</i>		1.58	1.42E-51	bifunctional folypolyglutamate synthase/dihydrofolate synthase
<i>JN057_RS07690</i>		1.77	7.86E-50	FAD-containing oxidoreductase
<i>JN057_RS01070</i>		1.54	8.80E-45	hypothetical protein
<i>JN057_RS08650</i>		1.70	1.72E-43	GlsB/YeaQ/YmgE family stress response membrane protein

(Table Continued)				
<i>JN057_RS06650</i>		1.66	1.77E-43	ABC transporter ATP-binding protein/permease
<i>JN057_RS08045</i>		1.51	1.24E-42	TIGR02206 family membrane protein
<i>JN057_RS04335</i>		1.69	1.70E-42	TIGR03943 family protein
<i>JN057_RS08280</i>		1.944	8.14E-40	ABC transporter ATP-binding protein
<b><i>JN057_RS08665</i></b>	<b><i>gls24</i></b>	1.63	2.81E-39	Asp23/Gls24 family envelope stress response protein
<i>JN057_RS04845</i>	<i>mscL</i>	1.77	6.12E-39	large conductance mechanosensitive channel protein MscL
<i>JN057_RS00735</i>		1.56	2.52E-35	rhodanese-related sulfurtransferase
<i>JN057_RS08655</i>	<i>amaP</i>	1.59	3.08E-35	alkaline shock response membrane anchor protein AmaP
<i>JN057_RS09005</i>		1.66	3.70E-35	ABC transporter permease
<i>JN057_RS04565</i>		1.68	6.80E-35	GNAT family N-acetyltransferase
<i>JN057_RS04255</i>		2.28	6.25E-33	ATP-grasp domain-containing protein
<i>JN057_RS06190</i>		1.56	1.25E-32	amino acid ABC transporter ATP-binding protein
<i>JN057_RS05915</i>	<i>nrdH</i>	1.71	1.55E-28	glutaredoxin-like protein NrdH
<i>JN057_RS03795</i>		1.57	3.47E-28	biotin transporter BioY
<i>JN057_RS09745</i>		2.58	8.36E-28	ABC transporter ATP-binding protein

**Table 6.5C:** Top 10 differentially downregulated early response genes by *S. pneumoniae* serotype 6B (BHN418) after 10 minutes incubation in iron-deplete THY (1mM BIP). Genes were ordered by statistical significance (P-adj). Top downregulated genes in both strains are in bold. RefSeq locus tags of the published sequence of BVJ1JL is shown (Betts *et al.* 2021).

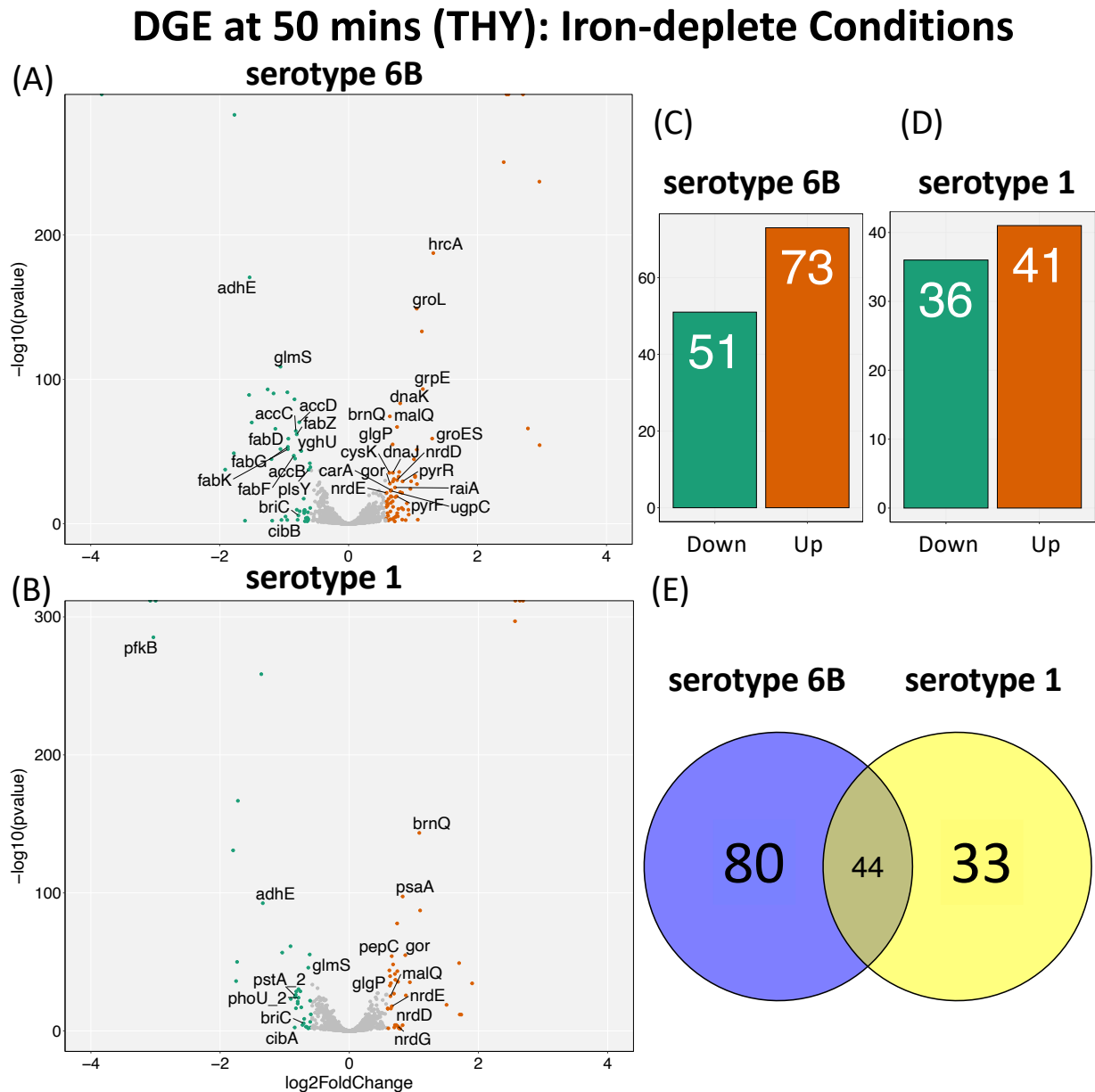
locus	Gene	FC	P-adj	Annotation
<b>Serotype 6B (BHN418)</b>				
<i>JN057_RS02195</i>		-4.02	1.33E-251	enoyl-CoA hydratase
<i>JN057_RS02215</i>	<i>fabK</i>	-3.04	3.21E-146	enoyl-[acyl-carrier-protein] reductase FabK
<i>JN057_RS02220</i>	<i>fabD</i>	-2.50	5.93E-115	ACP S-malonyltransferase
<i>JN057_RS03600</i>		-2.55	2.57E-107	DegV family protein
<i>JN057_RS09405</i>		-2.56	1.38E-102	DUF4231 domain-containing protein
<b><i>JN057_RS09915</i></b>	<b><i>adhE</i></b>	-2.20	5.57E-95	bifunctional acetaldehyde-CoA/alcohol dehydrogenase
<i>JN057_RS02225</i>	<i>fabG</i>	-2.34	4.34E-90	3-oxoacyl-[acyl-carrier-protein] reductase
<b><i>JN057_RS02620</i></b>	<b><i>dnaK</i></b>	-1.76	3.96E-80	molecular chaperone DnaK
<i>JN057_RS02230</i>	<i>fabF</i>	-2.13	2.97E-78	beta-ketoacyl-ACP synthase II
<i>JN057_RS02250</i>	<i>accD</i>	-1.73	4.89E-76	acetyl-CoA carboxylase, carboxyltransferase subunit beta

**Table 6.5D:** Top 10 differentially downregulated early response genes by *S. pneumoniae* serotype 1 (BVJ1JL) after 10 minutes incubation in iron-deplete THY (1mM BIP). Genes were ordered by statistical significance (P-adj). Top downregulated genes in both strains are in bold. RefSeq locus tags of the published sequence of BVJ1JL is shown (Betts *et al.* 2021).

locus	Gene	FC	P-adj	Annotation
<b>Serotype 1 (BVJ1JL)</b>				
<i>JN057_RS02625</i>		-4.21	2.86E-209	hypothetical protein
<i>JN057_RS02630</i>	<i>dnaJ</i>	-3.32	2.74E-200	molecular chaperone DnaJ
<b><i>JN057_RS02620</i></b>	<b><i>dnaK</i></b>	-3.53	5.94E-176	molecular chaperone DnaK
<i>JN057_RS09315</i>	<i>groL</i>	-2.50	3.66E-130	chaperonin GroEL
<i>JN057_RS06350</i>	<i>carB</i>	-1.92	1.88E-108	carbamoyl-phosphate synthase large subunit
<i>JN057_RS01605</i>	<i>manX</i>	-1.89	2.03E-108	PTS sugar transporter subunit IIB
<b><i>JN057_RS09915</i></b>	<b><i>adhE</i></b>	-2.56	2.38E-96	bifunctional acetaldehyde-CoA/alcohol dehydrogenase
<i>JN057_RS02610</i>	<i>hrcA</i>	-2.81	9.23E-94	heat-inducible transcriptional repressor HrcA
<i>JN057_RS10060</i>		-2.49	6.48E-88	zinc-dependent alcohol dehydrogenase family protein
<i>JN057_RS07005</i>		-2.02	2.38E-78	U32 family peptidase

### 6.3.9 Later transcriptomics response of *S. pneumoniae* during iron limitation in THY

In contrast to the data for 10 minutes incubation in iron-deplete THY, more genes were differentially expressed by serotype 6B (BHN418) after 50 minutes incubation in iron-deplete THY compared to the serotype 1 strain (**Figure 6.9**). Furthermore, there was a larger overlap in the number of DEGs between serotype 6B (BHN418) and serotype 1 (BVJ1JL) with 28.0% (44/157) of DEGs shared at 50 minutes (**Figure 6.9E**). All the genes upregulated by serotype 6B at 10 minutes increased expression further at the 50 minutes timepoint by magnitudes ranging from 33% to 276% (**Table 6.6A**). The genes with significantly upregulated expression by serotype 6B strain after 50 minutes of iron stress included the chaperon stress proteins (*hrcA*, *groELS*, *grpE*), pyrimidine metabolism genes (*pyrK*, *pyrDb*) and multiple ABC transporters including the virulence factor *aliB*, likely an adaptive mechanism to prolonged exposure to iron depletion (**Table 6.6A**). The gene expression of the *piu* (*piuABCD*) and *cop* (*copAY*, *cupA*) operons were statistically more significantly upregulated at 50 minutes compared to 10 minutes (**Table 6.6A** and **Table 6.6B**). The downregulated genes by serotype 6B were mainly involved in fatty acids and carbohydrate (PTS) transport and metabolism. The downregulated genes for serotype 1 were mostly related to PTS transport systems. For both serotype 1 and serotype 6B strains, *dpr* which is a pneumococcal iron storage protein, and *adhE*, an iron-binding alcohol dehydrogenase, were among the most significantly down regulated genes (**Table 6.6C** and **Table 6.6D**). *Dpr* and *adhE* are important for *S. pneumoniae* iron homeostasis and adaptation to oxidative stress (Andrews, Robinson, and Rodríguez-Quñones 2003; Echave *et al.* 2003; Ulijasz *et al.* 2004). The results of the iron limitation experiment in THY corroborated some previous findings relating to *S. pneumoniae* response to iron stress, with the expression of certain genes being correlated to iron availability. However, several strain-specific gene expression patterns were also observed here for clinical *S. pneumoniae* serotype 1 and serotype 6B strains in response to iron limitation.



**Figure 6.9: Differential gene expression (DGE) of BHN418 (serotype 6B) and BVJ1JL (serotype 1) after 50 minutes in iron-deplete environment in THY.** Volcano (A - B) and bar (C - D) plots showing up (brown) and down (green) regulated genes in serotype 6B and serotype 1 strains. Each point in the volcano plots indicate a differentially expressed gene (DEG) with colour indicative of expression direction (brown = upregulated, green = downregulated). Where available in the original annotation file, short gene names are annotated on volcano plots. (E) Venn diagram showing overlap in DEGs between serotype 6B and serotype 1 environment. DEGs were defined by fold change  $> 1.5$  and false discovery rate FDR  $< 0.05$ .

**Table 6.6A:** Top 35 differentially upregulated genes by *S. pneumoniae* serotype 6B (BHN418) after 50 minutes growth in iron-deplete THY (1mM BIP). Genes were ordered by statistical significance (P-adj). Top upregulated genes in both strains are in bold. RefSeq locus tags of the published sequence of BVJ1JL is shown (Betts *et al.* 2021).

locus	Gene	FC	P-adj	Annotation
<b>Serotype 6B (BHN418)</b>				
<b>JN057_RS09005</b>	<b><i>piuB</i></b>	6.52	0.00E+00	Iron compound ABC uptake transporter permease protein PiuB
<b>JN057_RS09010</b>	<b><i>piuC</i></b>	5.58	0.00E+00	Iron compound ABC uptake transporter permease protein PiuC
<b>JN057_RS09020</b>	<b><i>piuA</i></b>	5.46	0.00E+00	Iron compound ABC uptake transporter substrate-binding protein PiuA
<b>JN057_RS09015</b>	<b><i>piuD</i></b>	5.29	3.43E-251	Iron compound ABC uptake transporter ATP-binding protein PiuD
<b>JN057_RS03550</b>	<b><i>copA</i></b>	7.78	1.07E-237	heavy metal translocating P-type ATPase
<i>hrcA</i>	<i>hrcA</i>	2.49	3.36E-188	heat-inducible transcriptional repressor HrcA
<i>groL</i>	<u><i>groL</i></u>	2.08	1.03E-149	chaperonin GroEL
<i>JN057_RS03675</i>	<i>malT</i>	2.19	5.81E-134	PTS transporter subunit IIBC
<i>grpE</i>	<i>grpE</i>	2.22	5.15E-94	nucleotide exchange factor GrpE
<i>JN057_RS02620</i>	<i>dnaK</i>	1.74	4.01E-84	molecular chaperone DnaK
<i>JN057_RS03085</i>	<i>brnQ</i>	1.56	4.37E-75	branched-chain amino acid transport system II carrier protein
<i>JN057_RS10395</i>	<i>malQ</i>	1.68	9.10E-68	4-alpha-glucanotransferase
<b>JN057_RS03540</b>	<b><i>copY</i></b>	6.88	9.86E-67	CopY/TcrY family copper transport repressor
<i>groES</i>	<i>groES</i>	2.46	1.04E-59	co-chaperone GroES
<i>JN057_RS10390</i>	<i>glgP</i>	1.61	1.09E-55	glycogen/starch/alpha-glucan family phosphorylase
<b>JN057_RS03545</b>	<b><i>cupA</i></b>	7.79	3.63E-55	Copper-exporting ATPase

(Table Continued)

<b>JN057_RS07650</b>		2.08	4.67E-52	class I SAM-dependent methyltransferase
JN057_RS10400	<i>malX</i>	2.02	2.21E-45	extracellular solute-binding protein
JN057_RS10405		1.72	1.33E-36	ABC transporter permease subunit
JN057_RS02630	<i>dnaJ</i>	1.61	5.09E-36	molecular chaperone DnaJ
JN057_RS10875	<i>cysK</i>	1.55	5.59E-36	cysteine synthase A
JN057_RS08920	<i>paal</i>	2.04	6.80E-34	Paal family thioesterase
JN057_RS06400		1.71	2.41E-33	NCS2 family nucleobase:cation symporter
JN057_RS03680		2.04	4.77E-33	endonuclease/exonuclease/phosphatase family protein
JN057_RS06360		1.63	9.99E-32	aspartate carbamoyltransferase catalytic subunit
JN057_RS01205	<i>nrdD</i>	1.68	5.39E-31	anaerobic ribonucleoside-triphosphate reductase
JN057_RS04615	<i>pyrK</i>	1.96	3.92E-30	dihydroorotate dehydrogenase (NAD (+)), electron transfer subunit
JN057_RS06365	<i>pyrR</i>	1.79	4.41E-30	bifunctional pyr operon transcriptional regulator/uracil phosphoribosyltransferase PyrR
JN057_RS02985		1.61	4.51E-30	fructose-bisphosphate aldolase
JN057_RS03800	<i>gor</i>	1.56	1.58E-28	glutathione-disulfide reductase
<b>JN057_RS07405</b>	<b><i>aliB</i></b>	2.09	3.46E-28	peptide ABC transporter substrate-binding protein
JN057_RS10855	<i>raiA</i>	1.65	7.51E-26	ribosome-associated translation inhibitor
JN057_RS04620	<i>pyrDb</i>	1.94	6.80E-25	dihydroorotate dehydrogenase (NAD (+)), catalytic subunit
JN057_RS06355	<i>carA</i>	1.58	5.46E-24	carbamoyl phosphate synthase small subunit
JN057_RS07660	<i>ugpC</i>	1.57	1.43E-23	sn-glycerol-3-phosphate ABC transporter ATP-binding protein UgpC



**Table 6.6B:** Top 35 differentially upregulated genes by *S. pneumoniae* serotype 1 (BVJ1JL) after 50 minutes incubation in iron-deplete THY (1mM BIP). Genes were ordered by statistical significance (P-adj). Top upregulated genes in both strains are in bold. RefSeq locus tags of the published sequence of BVJ1JL are shown (Betts *et al.* 2021).

locus	Gene	FC	P-adj	Annotation
<b>Serotype 1 (BVJ1JL)</b>				
<b>JN057_RS09020</b>	<b><i>piuA</i></b>	6.45	0.00E+00	Iron compound ABC uptake transporter substrate-binding protein PiuA
<b>JN057_RS09005</b>	<b><i>piuB</i></b>	6.24	0.00E+00	Iron compound ABC uptake transporter permease protein PiuB
<b>JN057_RS09010</b>	<b><i>piuC</i></b>	5.94	0.00E+00	Iron compound ABC uptake transporter permease protein PiuC
<b>JN057_RS09015</b>	<b><i>piuD</i></b>	5.92	1.39E-297	Iron compound ABC uptake transporter ATP-binding protein PiuD
JN057_RS03085	<i>brnQ</i>	2.12	3.45E-144	branched-chain amino acid transport system II carrier protein
JN057_RS07965	<i>psaA</i>	1.77	4.03E-98	metal ABC transporter substrate-binding lipoprotein/adhesin PsaA
<b>JN057_RS07650</b>		2.14	5.44E-88	class I SAM-dependent methyltransferase
JN057_RS07960		1.67	1.35E-78	metal ABC transporter permease
JN057_RS03800	<i>gor</i>	1.83	1.32E-55	glutathione-disulfide reductase
JN057_RS01590	<i>pepC</i>	1.58	6.86E-55	aminopeptidase C
<b>JN057_RS03540</b>	<b><i>copY</i></b>	3.25	7.38E-50	CopY/TcrY family copper transport repressor
JN057_RS08980		1.60	6.87E-49	nuclear transport factor 2 family protein
JN057_RS08970		1.54	1.42E-44	ABC transporter permease/substrate-binding protein
JN057_RS07955		1.68	5.40E-44	metal ABC transporter ATP-binding protein
JN057_RS03675		1.63	5.35E-42	PTS transporter subunit IIBC
JN057_RS04800		1.55	2.22E-40	amino acid permease

(Table Continued)

<i>JN057_RS07690</i>		1.64	9.18E-38	FAD-containing oxidoreductase
<i>JN057_RS08920</i>		1.69	2.69E-36	Paal family thioesterase
<b><i>JN057_RS07405</i></b>	<b><i>aliB</i></b>	1.92	5.34E-36	peptide ABC transporter substrate-binding protein
<i>JN057_RS08975</i>		1.56	2.49E-35	ABC transporter ATP-binding protein
<b><i>JN057_RS03550</i></b>	<b><i>copA</i></b>	3.74	3.52E-35	heavy metal translocating P-type ATPase
<i>JN057_RS10390</i>	<i>glgP</i>	1.55	1.39E-33	glycogen/starch/alpha-glucan family phosphorylase
<i>JN057_RS03165</i>		1.62	1.00E-27	S8 family serine peptidase
<i>JN057_RS04230</i>		1.84	1.90E-26	hypothetical protein
<i>JN057_RS10395</i>	<i>malQ</i>	1.56	6.78E-26	4-alpha-glucanotransferase
<b><i>JN057_RS03545</i></b>	<b><i>cupA</i></b>	2.84	1.29E-19	Copper-exporting ATPase
<i>JN057_RS05920</i>	<i>nrdE</i>	1.58	1.37E-18	class 1b ribonucleoside-diphosphate reductase subunit alpha
<i>JN057_RS03970</i>		1.51	4.76E-17	amino acid ABC transporter ATP-binding protein
<i>JN057_RS06775</i>		1.56	6.27E-17	alpha-amylase
<i>JN057_RS08985</i>		1.51	9.53E-17	MarR family transcriptional regulator
<b><i>JN057_RS00825</i></b>		3.32	1.58E-12	ABC transporter substrate-binding protein
<b><i>JN057_RS00830</i></b>	<b><i>argG</i></b>	3.28	1.36E-12	argininosuccinate synthase
<i>JN057_RS01205</i>	<i>nrdD</i>	1.66	5.69E-05	anaerobic ribonucleoside-triphosphate reductase
<i>group_418</i>		1.63	5.79E-05	unannotated protein
<i>JN057_RS01225</i>		1.77	7.22E-05	phosphoribulokinase

**Table 6.6C:** Top 10 differentially downregulated genes by *S. pneumoniae* serotype 6B (BHN418) after 50 minutes incubation in iron-deplete THY (1mM BIP). Genes were ordered by statistical significance (P-adj). Bold gene names are among the top downregulated genes in both BHN418 and BVJ1JL. RefSeq locus tags of the published sequence of BVJ1JL are shown (Betts *et al.* 2021).

locus	Gene	FC	padj	Annotation
<b>Serotype 6B (BHN418)</b>				
<b>JN057_RS04210</b>	<b>fruR</b>	-16.98	0.00E+00	Transcriptional repressor of the fructose operon, DeoR family
<b>JN057_RS04215</b>	<b>fruB</b>	-16.18	0.00E+00	1-phosphofructokinase
<b>JN057_RS04220</b>	<b>fruA</b>	-14.22	0.00E+00	fructose-specific PTS transporter subunit EIIC
<b>JN057_RS07620</b>	<b>dpr</b>	-3.41	6.01E-284	DNA starvation/stationary phase protection protein/ Bifunctional non-specific DNA-binding protein/ferroxidase Dpr
<b>JN057_RS09915</b>	<b>adhE</b>	-2.90	2.18E-171	bifunctional acetaldehyde-CoA/alcohol dehydrogenase
<b>JN057_RS01515</b>	<b>glmS</b>	-2.08	1.65E-109	glutamine--fructose-6-phosphate transaminase (isomerizing)
<b>JN057_RS03600</b>		-2.39	8.54E-94	DegV family protein
<b>JN057_RS02205</b>		-1.93	8.63E-92	ketoacyl-ACP synthase III
<b>JN057_RS02195</b>		-2.24	5.50E-91	enoyl-CoA hydratase
<b>JN057_RS09895</b>	<b>bguA</b>	<b>-2.91</b>	6.20E-90	glycoside hydrolase family 1 protein

**Table 6.6D:** Top 10 differentially downregulated genes by *S. pneumoniae* serotype 1 (BVJ1JL) after 50 minutes incubation in iron-deplete THY (1mM BIP). Genes were ordered by statistical significance (P-adj). Bold gene names are among the top downregulated genes in both BHN418 and BVJ1JL. RefSeq locus tags of the published sequence of BVJ1JL are shown (Betts *et al.* 2021).

locus	Gene	FC	P-adj	Annotation
<b>Serotype 1 (BVJ1JL)</b>				
<b>JN057_RS04210</b>	<b>fruR</b>	-8.47	0.00E+00	Transcriptional repressor of the fructose operon, DeoR family
<b>JN057_RS04220</b>	<b>fruA</b>	-7.98	0.00E+00	fructose-specific PTS transporter subunit EIIC
<b>JN057_RS04215</b>	<b>fruB</b>	-8.18	6.03E-286	1-phosphofructokinase
<b>JN057_RS07620</b>	<b>dpr</b>	-2.57	3.13E-259	DNA starvation/stationary phase protection protein/ Bifunctional non-specific DNA-binding protein/ferroxidase Dpr
<b>JN057_RS09895</b>	<b>bguA</b>	-3.30	1.87E-167	6-phospho-beta-glycosidase
<b>JN057_RS09900</b>	<b>bguD</b>	-3.48	1.52E-131	PTS sugar transporter subunit IIC
<b>JN057_RS09915</b>	<b>adhE</b>	-2.53	2.78E-93	bifunctional acetaldehyde-CoA/alcohol dehydrogenase
JN057_RS07005		-1.88	4.44E-62	U32 family peptidase
group_427		-2.05	1.89E-57	unannotated protein
<b>JN057_RS02205</b>		-1.53	4.28E-56	ketoacyl-ACP synthase III

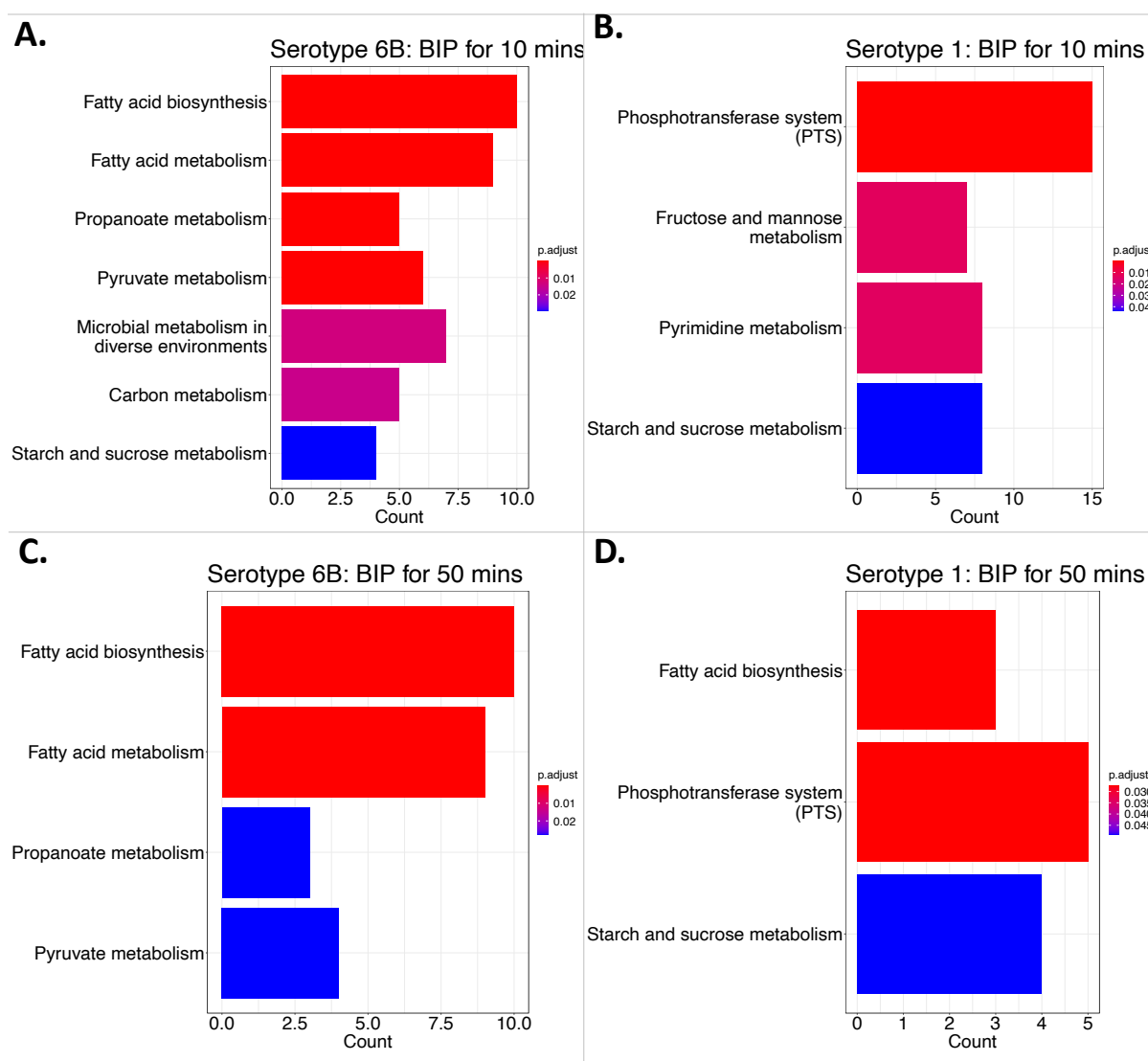
### 6.3.10 Enriched biological pathways by *S. pneumoniae* in iron-deplete conditions in THY

Using *S. pneumoniae* KEGG database annotations for the serotype 6B strain 670-6B and the serotype 1 strain INV104, we interrogated which biological pathways were enriched during iron-depletion in THY. These reference strains were chosen because they were among the few *S. pneumoniae* strains with annotated biological pathways available on KEGG and are close representatives of our study strains. However, some of the genes that were differentially expressed in our study strains (BHN418 and BVJ1JL) were not present/annotated in these reference strains (**Table 6.12**). The early response genes upregulated by both serotype 1 (BVJ1JL) and serotype 6B (BHN418) at 10 minutes in iron-deplete THY were not statistically enriched for any biological pathways. Following 50 minutes exposure to iron limitation in THY, the upregulated genes of serotype 6B (BHN418) strain were significantly enriched for pyrimidine metabolism while serotype 1 (BVJ1JL) was most significantly enriched for ABC transporter pathway genes (**Table 6.7**). The downregulated pathways that were enriched by both strains at early and later exposure to iron-depletion in THY were mainly involved with fatty-acid biosynthesis, amino acid biosynthesis, and carbohydrate export and metabolism (**Figure 6.10**).

**Table 6.7:** Pathways upregulated by of *S. pneumoniae* serotype 1 (BVJ1JL) and serotype 6B (BHN418) after 50 minutes incubation in iron-deplete conditions in THY. Orthologs of KEGG annotations for reference strains INV104 (serotype 1) and 670-6B (serotype 6B) were used where available.

Pathway ID	Description	GeneRatio	BgRatio	p.adjust	qvalue
<b>Upregulated pathways by Serotype 6B (BHN418), 50 mins in iron-deplete THY</b>					
snb00240	Pyrimidine metabolism	10/27	30/746	2.38E-07	1.88E-07
<b>Upregulated pathways by Serotype 1 (BVJ1JL), 50 mins in iron-deplete THY</b>					
sni02010	ABC transporters	7/15	98/732	0.02	0.01
sni00500	Starch and sucrose metabolism	4/15	37/732	0.03	0.02

## Significantly Downregulated Pathways : THY



**Figure 6.10: Significantly downregulated pathways by *S. pneumoniae* serotype 6B (BHN418) and Serotype 1 (BVJ1JL) strains after 10 minutes and 50 minutes exposure to iron-limitation in THY.** Orthologs present in KEGG annotated strain for 670-6B (serotype 6B) and INV104 (serotype 1) were used. Pathway enrichment was performed by over representation analysis and pathways with FDR <0.05 were considered significantly enriched. Only pathways with FDR <0.05 are shown. Length of the bar corresponds to the number of DEGs in respective pathways and colour indicates significance level with red being most statistically significant and blue being least significant for each plot.

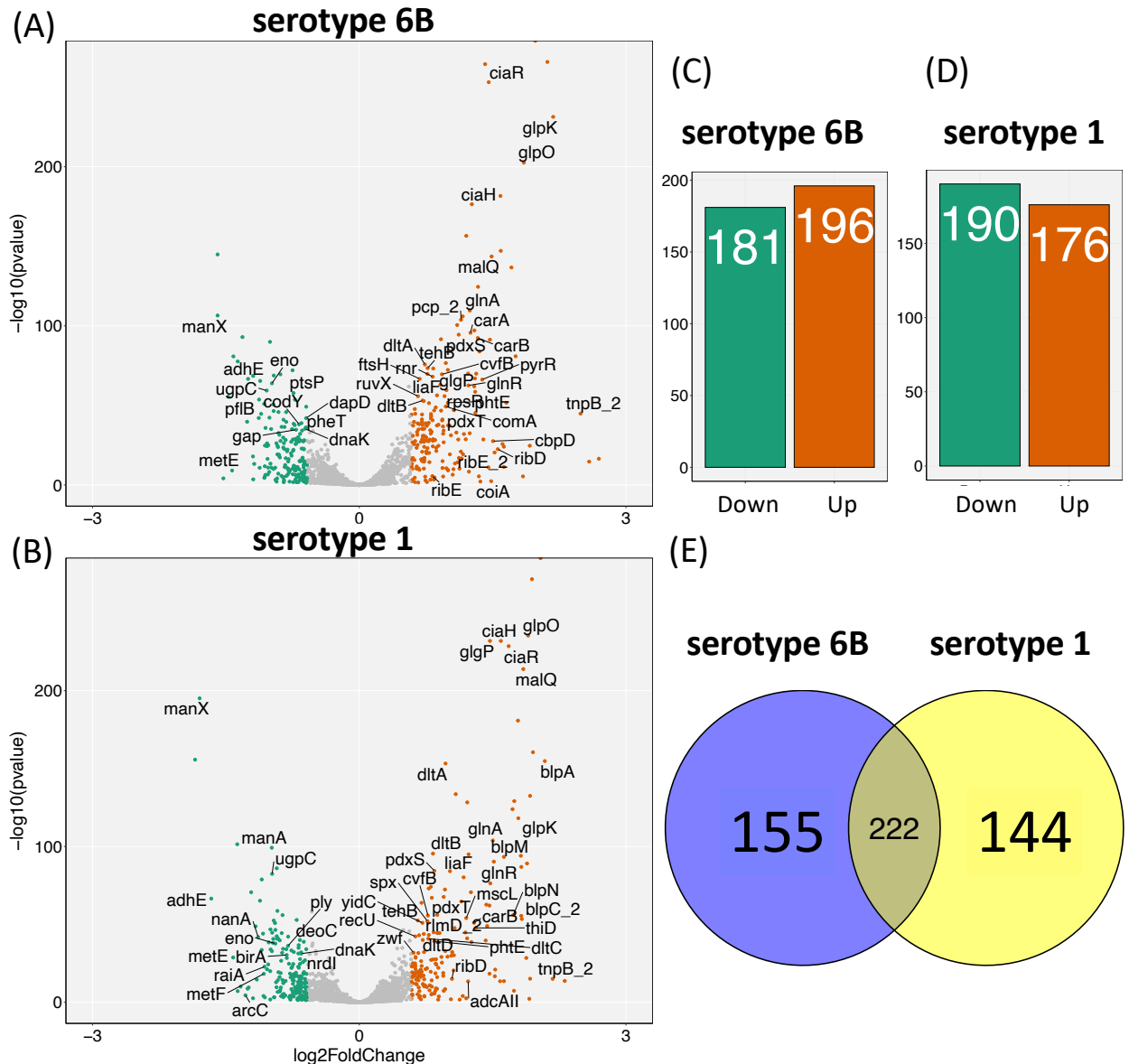
### 6.3.11 Early transcriptomic response of *S. pneumoniae* during iron limitation in CDM

To explore the effects of cation stress further, RNA-seq experiments were repeated using iron depletion and repletion in CDM media where the concentration of iron availability is more controlled. In contrast to growth in iron depleted THY, induction of low iron levels in CDM resulted in many more DEGs for both serotype 6B (BHN418) and serotype 1 (BVJ1JL), implying a more stress growth environment (**Figure 6.11**). The top three most significantly upregulate genes by serotype 6B, and the top two by serotype 1 were proteins of undefined functions (**Table 6.8A** and **Table 6.8B**). One of these genes, *JN057\_RS03040* (*SP\_1027* in TIGR4) was absent in D39V strain and is annotated as a putative metalloprotease in other strains, suggesting a role in adaptation to metals stress. The most significantly upregulated gene by serotype 1 and the third most upregulated by serotype 6B, *JN057\_RS04930*, was an extracellular protein which was annotated as a domain of unknown function (DUF) in all reference strains. Another extracellular-like protein significantly upregulated by serotype 6B, *JN057\_RS04960*, was annotated in D39V as a putative phage shock protein C but described in the other reference strains used as a *pspC* containing domain protein. The two-component system sensor histidine kinase *ciaH*, and its cognate response regulator *ciaR* were among the most significantly upregulated genes by both strains (**Table 6.8A** and **Table 6.8B**). The serotype 6B (BHN418) strain upregulated several regulators such as *pyrR*, *pmrA*, *cvfB*, *JN057\_RS10095*, pyrimidine (*pyrK*, *pyrR*, *pyrB*, *pyrD*), and amino acid (*carAB*) biosynthesis genes. The iron storage and homeostasis protein, *dpr*, and a mercury reductase gene *merA* were also upregulated by serotype 6B. The serotype 1 (BVJ1JL) strain upregulated several ABC transporters (*glnP2a*, *glnP2b*, *glnH2*, *JN057\_RS03005*, *JN057\_RS03010*) and more two-component sensor systems (*vraR*, *vraT*). Some of the bacteriocin like peptides (*blpA*, *blpM*, *blpT*, *blpN*, *blpC*) genes were uniquely upregulated by serotype 1 in early adaptation to iron-deplete CDM, as well as the lipoteichoic acid biosynthesis gene *dltA*. *S. pneumoniae* DNA processing protein *dprA*, which is expressed under conditions that induce DNA competence (Campbell, Choi, and Masure 1998), was significant upregulated by both serotype 1 (BVJ1JL) and serotype 6B (BHN418) at this timepoint in iron-deplete CDM. *brnQ*, *piu* and *cop* operon genes, which were significantly upregulated early and late in iron-deplete THY by both serotype 1 and serotype 6B, were not differentially expressed in iron-deplete CDM. The oxidative stress response gene *adhE* was downregulated similar to iron-deplete

condition in THY. In both strains, the downregulated genes were dominated by carbohydrate metabolism (PTS) genes. The pneumococcal virulence gene neuraminidase A (*nanA*) was also among the downregulated genes by both strains.



## DGE at 10 mins (CDM): Iron-deplete Conditions



**Figure 6.11: Differential gene expression (DGE) of BHN418 (serotype 6B) and BVJ1JL (serotype 1) after 10 minutes in iron-deplete condition (1mM BIP) in CDM.** Volcano (A - B) and bar (C -D) plots showing up (brown) and down (green) regulated genes in serotype 6B and serotype 1 strains. Each point in the volcano plots indicate a gene expression with colour indicative of expression direction (brown = upregulated, green = downregulated). Where available in the original file, gene names are annotated on volcano plots, allowing maximum of 40 gene annotations on each plot. **(E)** Venn diagram showing overlap in DEGs between serotype 6B and serotype 1 at this timepoint. DEGs were defined by fold change > 1.5 and false discovery rate FDR < 0.05.

**Table 6.8A:** The top 35 differentially upregulated genes by *S. pneumoniae* serotype 6B (BHN418) after 10 minutes incubation in iron-deplete CDM (1mM BIP). Genes were ranked by statistical significance (P-adj). RefSeq locus tags of the published sequence of BVJ1JL are shown.

locus	Gene	FC	P-adj	Annotation
<b>Serotype 6B (BHN418)</b>				
<i>JN057_RS03040</i>		3.95	0.00E+00	M57 family metalloprotease
<i>JN057_RS04360</i>		4.33	1.50E-266	PspC domain-containing protein
<i>JN057_RS04930</i>		2.67	3.63E-265	DUF1002 domain-containing protein
<i>JN057_RS03870</i>	<i>ciaR</i>	2.74	7.77E-254	two-component system response regulator CiaR
<i>JN057_RS10760</i>	<i>glpK</i>	4.54	4.92E-232	glycerol kinase GlpK
<i>JN057_RS10755</i>	<i>glpO</i>	3.61	2.78E-203	type 1 glycerol-3-phosphate oxidase
<i>JN057_RS01185</i>	<i>folC</i>	3.01	2.33E-182	bifunctional folylpolyglutamate synthase/dihydrofolate synthase
<i>JN057_RS03875</i>	<i>ciaH</i>	2.41	3.81E-177	two-component system sensor histidine kinase CiaH
<i>JN057_RS11045</i>	<i>htrA</i>	2.30	2.84E-157	Serine protease, DegP/HtrA
<i>JN057_RS10750</i>	<i>glpF</i>	3.01	7.25E-148	Glycerol uptake facilitator protein
<i>JN057_RS10395</i>	<i>malQ</i>	2.81	2.56E-144	4-alpha-glucanotransferase
<i>JN057_RS10090</i>		3.28	2.17E-137	DUF1304 domain-containing protein
<i>JN057_RS01190</i>		2.52	2.94E-125	SP_0198 family lipoprotein
<i>JN057_RS02550</i>	<i>glnA</i>	2.37	2.48E-110	type I glutamate--ammonia ligase
<i>JN057_RS07620</i>	<i>dpr</i>	2.24	1.08E-106	Bifunctional non-specific DNA-binding protein/ferroxidase Dpr
<i>pcp_2</i>	<i>pcp_2</i>	2.21	1.62E-104	Pyrrolidone-carboxylate peptidase, degenerate (pseudogene)
<i>JN057_RS04375</i>		2.15	3.10E-101	ABC transporter permease
<i>JN057_RS01340</i>	<i>adk</i>	2.45	7.54E-98	adenylate kinase

(Table Continued)

JN057_RS06355	<i>carA</i>	2.38	2.34E-96	carbamoyl phosphate synthase small subunit
JN057_RS07130	<i>pdxS</i>	2.17	3.35E-95	pyridoxal 5'-phosphate synthase lyase subunit PdxS
JN057_RS06350	<i>carB</i>	2.52	4.17E-93	carbamoyl-phosphate synthase large subunit
JN057_RS07690	<i>merA</i>	1.89	2.59E-92	Pyridine nucleotide-disulfide oxidoreductase
JN057_RS04620	<i>pyrDb</i>	2.77	4.93E-92	dihydroorotate dehydrogenase (NAD (+)), catalytic subunit
JN057_RS06775	<i>amy</i>	2.55	1.29E-84	Cytoplasmic alpha-amylase
JN057_RS10095		3.38	1.30E-81	MarR family winged helix-turn-helix transcriptional regulator
JN057_RS04660	<i>pmrA</i>	1.96	2.28E-77	multidrug efflux MFS transporter PmrA
JN057_RS10725	<i>dltA</i>	1.67	1.96E-76	D-alanine--poly(phosphoribitol) ligase subunit DltA
JN057_RS04685	<i>tehB</i>	1.70	3.71E-74	SAM-dependent methyltransferase TehB
JN057_RS07105		1.78	6.43E-74	methylated-DNA--[protein]-cysteine S-methyltransferase
JN057_RS11050	<i>parB</i>	1.99	4.72E-73	ParB/RepB/Spo0J chromosome family partition protein
JN057_RS07680	<i>cshA</i>	2.34	9.30E-71	DEAD/DEAH box helicase
JN057_RS06360	<i>pyrB</i>	2.49	1.11E-70	aspartate carbamoyltransferase catalytic subunit
JN057_RS06315		1.70	1.57E-70	DMT family transporter
JN057_RS04535	<i>cvfB</i>	1.90	2.86E-70	RNA-binding virulence regulatory protein CvfB
JN057_RS04675	<i>rnr</i>	1.77	9.59E-69	ribonuclease R

**Table 6.8B:** The top 35 differentially upregulated genes by *S. pneumoniae* serotype 1 (BVJ1JL) after 10 minutes incubation in iron-deplete CDM (1mM BIP). Genes were ranked by statistical significance (P-adj). Bold gene names are upregulated in both serotype 1 and serotype 6B. RefSeq locus tags of the published sequence of BVJ1JL are shown.

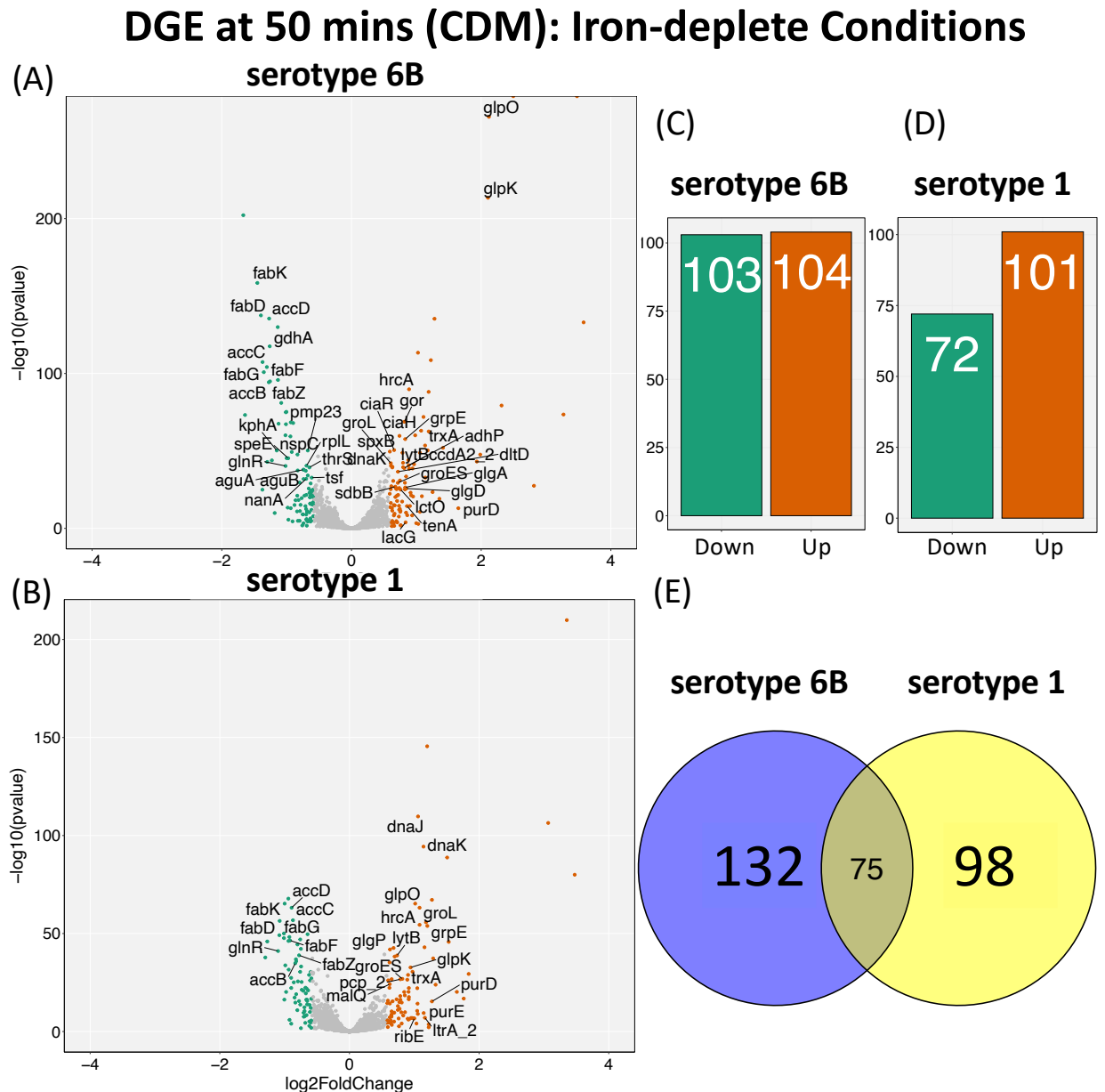
locus	Gene	FC	P-adj	Annotation
Serotype 1 (BVJ1JL)				
<b>JN057_RS04930</b>		4.11	0.00E+00	DUF1002 domain-containing protein
<b>JN057_RS03040</b>		3.84	3.06E-272	M57 family metalloprotease
<b>JN057_RS10755</b>	<b>glpO</b>	3.73	4.74E-236	type 1 glycerol-3-phosphate oxidase
<b>JN057_RS10390</b>	<b>glgP</b>	2.77	1.26E-232	glycogen/starch/alpha-glucan family phosphorylase
<b>JN057_RS03875</b>	<b>ciaH</b>	3.02	1.29E-232	two-component system sensor histidine kinase CiaH
<b>JN057_RS03870</b>	<b>ciaR</b>	3.21	2.90E-229	two-component system response regulator CiaR
<b>JN057_RS10395</b>	<b>malQ</b>	3.59	1.52E-214	4-alpha-glucanotransferase
<b>JN057_RS03005</b>	<b>glnH2</b>	3.45	1.79E-181	Glutamine ABC transporter substrate-binding protein
<b>JN057_RS04360</b>		3.87	3.31E-161	PspC domain-containing protein
<b>JN057_RS02690</b>	<b>blpA</b>	4.25	1.61E-155	peptide cleavage/export ABC transporter BlpA
<b>JN057_RS10725</b>	<b>dltA</b>	1.96	6.12E-154	D-alanine--poly(phosphoribitol) ligase subunit DltA
<b>JN057_RS09585</b>		2.12	2.78E-134	Bax inhibitor-1/YccA family protein
<b>JN057_RS03000</b>	<b>glnP2a</b>	3.79	3.60E-133	Glutamine ABC transporter permease
<b>JN057_RS10750</b>	<b>glpF</b>	3.35	7.92E-130	Glycerol uptake facilitator protein
<b>JN057_RS07680</b>	<b>cshA</b>	2.32	5.14E-129	DEAD/DEAH box helicase
<b>JN057_RS11045</b>	<b>htrA</b>	3.30	1.23E-124	Serine protease, DegP/HtrA
<b>JN057_RS10760</b>	<b>glpK</b>	3.46	7.74E-119	glycerol kinase GlpK
<b>JN057_RS02550</b>	<b>glnA</b>	2.83	1.49E-104	type I glutamate--ammonia ligase

(Table Continued)

<i>JN057_RS10720</i>	<i>dltB</i>	1.78	4.08E-96	D-alanyl-lipoteichoic acid biosynthesis protein DltB
<i>JN057_RS01185</i>	<i>folC</i>	2.35	1.21E-95	bifunctional folylpolyglutamate synthase/dihydrofolate synthase
<i>JN057_RS02995</i>	<i>glnP2b</i>	3.52	1.00E-94	Glutamine ABC transporter permease
<i>JN057_RS02695</i>	<i>blpM</i>	3.09	5.00E-94	two-peptide bacteriocin subunit BlpM
<i>JN057_RS04375</i>		2.85	5.861E-91	ABC transporter permease
<i>JN057_RS03010</i>		3.70	8.56E-90	amino acid ABC transporter ATP-binding protein
<i>JN057_RS02685</i>	<i>blpB</i>	3.54	1.39E-87	bacteriocin secretion accessory protein
<i>JN057_RS07130</i>	<i>pdxS</i>	1.80	4.60E-85	pyridoxal 5'-phosphate synthase lyase subunit PdxS
<i>liaF</i>	<i>liaF/vra</i>	2.03	7.46E-85	cell wall-active antibiotics response protein
<i>JN057_RS02085</i>	<i>T</i>			
<i>JN057_RS01190</i>		2.26	6.95E-81	SP_0198 family lipoprotein
<i>JN057_RS02545</i>	<i>glnR</i>	2.78	5.05E-77	transcriptional repressor GlnR
<i>JN057_RS04610</i>		1.75	7.46E-75	VOC family protein
<i>JN057_RS06325</i>		1.72	9.76E-74	alcohol dehydrogenase catalytic domain-containing protein
<i>JN057_RS02095</i>	<i>vraR</i>	1.95	4.37E-73	Two-component transcriptional regulator vraR
<i>JN057_RS11050</i>	<i>parB</i>	2.39	2.49E-71	ParB/RepB/Spo0J family partition protein
<i>JN057_RS01340</i>	<i>adk</i>	1.93	2.39E-68	adenylate kinase
<i>JN057_RS04340</i>		1.76	9.04E-68	permease

### 6.3.12 Later transcriptomic response of *S. pneumoniae* during iron limitation in CDM

The gene expression profile of serotype 1 and serotype 6B was dramatically different after 50 minutes in iron deplete conditions in CDM with lower number of DEGs compared to the 10 minutes timepoint (**Figure 6.12**). At 50 minutes, more pneumococcal genes specifically associated with transition metals homeostasis and several ABC transporters were upregulated for both strains (**Table 6.9A** and **Table 6.9B**). In both the serotype 6B (BHN418) and serotype 1 (BVJ1JL) strains, the copper efflux system genes (*copA*, *copY*, *cupA*) were the most significantly upregulated genes by effect size (fold change) (**Table 6.9A** and **Table 6.9B**). The *cop* operon genes were not differentially expressed at earlier timepoint of 10 minutes by either strain in iron deplete CDM. The enhanced gene expression of *cop* operon at 50 minutes was also observed in THY media, suggesting that the operon may be important for *S. pneumoniae* adaptation to prolonged iron limitation. However, the *piu* operon was not significantly upregulated in iron depleted CDM at 50 minutes. Significant upregulation of several ABC transporters for both strains emphasised the importance of substrates uptake/export for *S. pneumoniae* adaptation to low iron (**Table 6.9A** and **Table 6.9B**). In addition to several upregulated hypothetical proteins of unknown function (*SP\_2183 (TIGR4)*, *JN057\_RS08475*, *JN057\_RS00810*, *JN057\_RS00810*, *JN057\_RS07420*, *JN057\_RS08670*, *JN057\_RS01090*), many of the ABC transporters (*JN057\_RS04370*, *JN057\_RS04375*, *JN057\_RS04335*) upregulated at this time point have undefined functions, raising the possibility that these genes may be components of the *S. pneumoniae* iron response repertoire (**Table 6.9A** and **Table 6.9B**). Upregulation of chaperone stress response proteins (*dnaJ*, *dnaK*, *groL*, *grpE*, *clpL*, *hrcA*) was a feature of BVJ1JL (serotype 1) after 50 minutes exposure to iron depleted CDM. Fatty acid synthesis genes, some ABC transporters, and sugar metabolism genes were among the downregulated genes by both strains at this condition.



**Figure 6.12: Differential gene expression (DGE) of BHN418 (serotype 6B) and BVJ1JL (serotype 1) after 50 minutes in iron-deplete CDM.** Volcano (A - B) and bar (C - D) plots showing up (brown) and down (green) regulated genes in serotype 6B and serotype 1 strains. Each point in the volcano plots indicate gene expression with colour indicative of expression direction (brown = upregulated, green = downregulated). Where available in the original file, gene names were annotated on volcano plots, allowing maximum of 40 gene annotations on each plot. (E) Venn diagram showing overlap in DEGs between serotype 6B and serotype 1 at this timepoint. DEGs were defined by fold change >1.5 and false discovery rate FDR <0.05.

**Table 6.9A:** The top 35 differentially upregulated genes by *S. pneumoniae* serotype 6B (BHN418) after 50 minutes incubation in iron-deplete CDM (1mM BIP). Genes were ranked by statistical significance (P-adj). Top upregulated genes by both serotype 1 and serotype 6B are in bold. RefSeq locus tags of the published sequence of BVJ1JL are shown.

locus	Gene	FC	P-adj	Annotation
<b>Serotype 6B (BHN418)</b>				
<b>JN057_RS03550</b>	<b>copA</b>	11.14	0.00E+00	heavy metal translocating P-type ATPase
<b>JN057_RS10750</b>	<b>glpF</b>	5.64	0.00E+00	Glycerol uptake facilitator protein
<b>JN057_RS10755</b>	<b>glpO</b>	4.35	1.84E-266	type 1 glycerol-3-phosphate oxidase
JN057_RS10760	glpK	4.30	3.42E-214	glycerol kinase GlpK
<b>JN057_RS04375</b>		2.43	4.50E-136	ABC transporter permease
JN057_RS03540	copY	11.96	9.85E-134	CopY/TcrY family copper transport repressor
<b>JN057_RS07690</b>	<b>merA</b>	2.04	3.76E-114	Pyridine nucleotide-disulfide oxidoreductase
JN057_RS10820	thiX	2.34	2.45E-109	Thiamine ABC transporter permease
<b>JN057_RS02610</b>	<b>hrcA</b>	1.85	1.54E-90	heat-inducible transcriptional repressor HrcA
JN057_RS10810	thiZ	2.28	7.36E-89	Thiamine ABC transporter ATP-binding protein
SP_2183 (TIGR4)		4.97	4.60E-80	hypothetical protein
<b>JN057_RS03545</b>	<b>cupA</b>	9.64	2.98E-74	Copper-exporting ATPase
JN057_RS10815	thiY	2.16	1.12E-72	Thiamine ABC transporter substrate-binding protein
JN057_RS03800	gor	1.76	1.40E-69	glutathione-disulfide reductase
<b>JN057_RS01620</b>		2.11	8.59E-64	Guanine-hypoxanthine permease
<b>JN057_RS08475</b>		2.28	2.71E-63	DUF4649 family protein
<b>JN057_RS08970</b>	<b>proWX</b>	1.98	8.22E-61	ABC transporter permease/substrate-binding protein
JN057_RS03875	ciaH	1.67	1.82E-60	two-component system sensor histidine kinase CiaH



(Table Continued)

<i>JN057_RS02615</i>	<i>grpE</i>	1.78	2.27E-58	nucleotide exchange factor GrpE
<b><i>JN057_RS08480</i></b>	<b><i>trxA</i></b>	2.20	3.08E-54	thioredoxin
<b><i>JN057_RS04370</i></b>		2.66	9.43E-53	ABC transporter ATP-binding protein
<i>JN057_RS03870</i>	<i>ciaR</i>	1.58	2.71E-51	two-component system response regulator CiaR
<i>JN057_RS07685</i>	<i>oxIT</i>	2.26	4.64E-51	Major facilitator oxalate:formate antiporter
<i>JN057_RS03555</i>	<i>spxB</i>	1.51	2.51E-50	pyruvate oxidase
<i>JN057_RS04340</i>		1.73	1.54E-49	permease
<i>JN057_RS00465</i>		2.08	1.68E-49	CoA-binding protein
<i>JN057_RS10630</i>	<i>adh2</i>	3.96	2.19E-48	iron-containing alcohol dehydrogenase
<b><i>JN057_RS08975</i></b>	<b><i>proV</i></b>	1.84	9.77E-45	Glycine betaine/L-proline ABC transporter ATP-binding protein
<b>group_1142</b>		3.82	6.99E-44	unannotated protein
<i>JN057_RS04625</i>	<i>lytB</i>	1.73	4.81E-43	glucosaminidase domain-containing protein
<i>JN057_RS09315</i>	<i>groL</i>	1.52	9.03E-43	chaperonin GroEL
<i>JN057_RS00810</i>		1.95	3.98E-42	Putative bacteriocin
<i>JN057_RS01610</i>	<i>adhA</i>	1.84	1.85E-41	alcohol dehydrogenase
<i>JN057_RS11045</i>		1.53	4.22E-41	trypsin-like peptidase domain-containing protein
<i>JN057_RS02620</i>	<i>dnaK</i>	1.55	1.98E-40	molecular chaperone DnaK

**Table 6-9B:** The top 35 differentially upregulated genes by *S. pneumoniae* serotype 1 (BVJ1JL) after 50 minutes incubation in iron-deplete CDM (1mM BIP). Genes were ranked by statistical significance (P-adj). Top upregulated genes by both serotype 1 and serotype 6B are in bold. RefSeq locus tags of the published sequence of BVJ1JL are shown.

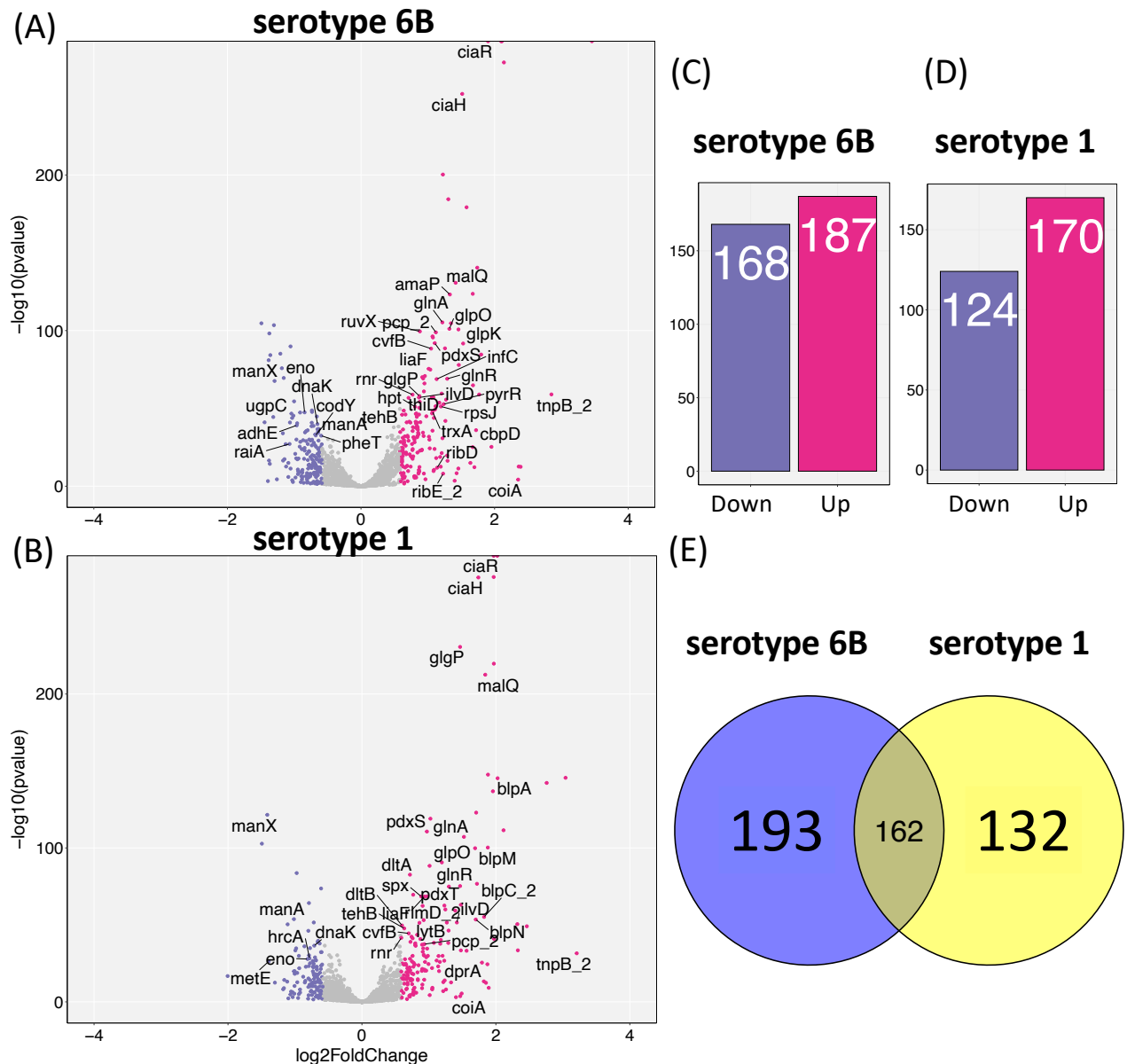
locus	Gene	FC	P-adj	Annotation
<b>Serotype 1 (BVJ1JL)</b>				
<b>JN057_RS03550</b>	<b>copA</b>	10.22	1.25E-210	heavy metal translocating P-type ATPase
<b>JN057_RS01620</b>		2.30	2.90E-146	Guanine-hypoxanthine permease
JN057_RS02630	dnaJ	2.08	1.82E-110	molecular chaperone DnaJ
<b>JN057_RS03540</b>	<b>copY</b>	8.37	3.82E-107	CopY/TcrY family copper transport repressor
JN057_RS02620	dnaK	2.21	4.56E-95	molecular chaperone DnaK
<b>JN057_RS04375</b>		2.84	1.63E-89	ABC transporter permease
<b>JN057_RS03545</b>	<b>cupA</b>	11.11	9.80E-81	cupredoxin domain-containing protein
<b>JN057_RS10750</b>	<b>glpF</b>	2.42	6.52E-68	Glycerol uptake facilitator protein
<b>JN057_RS10755</b>	<b>glpO</b>	2.02	5.50E-66	type 1 glycerol-3-phosphate oxidase
JN057_RS09315	groL	2.12	6.88E-64	chaperonin GroEL
JN057_RS00800		2.27	1.88E-56	LysM peptidoglycan-binding domain-containing protein
<b>JN057_RS02610</b>	<b>hrcA</b>	2.12	3.57E-55	heat-inducible transcriptional repressor HrcA
JN057_RS02615	grpE	2.30	1.32E-54	nucleotide exchange factor GrpE
<b>JN057_RS04370</b>		2.89	1.85E-46	ABC transporter ATP-binding protein
JN057_RS01840	clpL	2.23	1.03E-43	ATP-dependent Clp protease ATP-binding subunit
JN057_RS04930		1.60	2.31E-43	DUF1002 domain-containing protein
JN057_RS10390	glgP	1.54	1.30E-42	glycogen/starch/alpha-glucan family phosphorylase
JN057_RS04625	lytB	1.66	1.50E-39	glucosaminidase domain-containing protein
JN057_RS06130	fhs	1.62	5.48E-39	formate--tetrahydrofolate ligase
(Table Continued)				

<i>JN057_RS10100</i>		2.45	4.73E-38	LysM peptidoglycan-binding domain-containing protein
<i>JN057_RS01575</i>		1.54	5.54E-36	aminopeptidase
<i>JN057_RS10760</i>	<i>glpK</i>	1.92	2.05E-33	glycerol kinase GlpK
<i>JN057_RS10095</i>		1.96	3.19E-31	MarR family winged helix-turn-helix transcriptional regulator
<b>group_1142</b>		3.58	4.12E-30	unannotated protein
<i>JN057_RS08970</i>	<i>proWX</i>	1.87	1.31E-29	ABC transporter permease/substrate-binding protein
<i>JN057_RS09320</i>	<i>groES</i>	1.76	1.36E-27	co-chaperone GroES
<i>JN057_RS00785</i>		1.57	1.90E-27	HAD family hydrolase
<i>pcp_2</i>		1.72	2.10E-27	
<i>JN057_RS07690</i>	<i>merA</i>	1.86	3.59E-27	Pyridine nucleotide-disulfide oxidoreductase
<i>JN057_RS10905</i>		1.51	7.69E-27	CHAP domain-containing protein
<i>JN057_RS10395</i>	<i>malQ</i>	1.54	6.80E-25	4-alpha-glucanotransferase
<i>JN057_RS00525</i>	<i>purL</i>	2.51	1.27E-24	phosphoribosylformylglycinamide synthase
<i>JN057_RS01190</i>		1.53	4.93E-23	SP_0198 family lipoprotein
<i>JN057_RS08980</i>		1.89	6.41E-23	nuclear transport factor 2 family protein
<i>JN057_RS08480</i>	<i>trxA</i>	2.07	7.76E-23	thioredoxin

### 6.3.13 Early transcriptomics response of *S. pneumoniae* in iron replete CDM

Exposure of *S. pneumoniae* to iron-replete conditions showed broad transcriptomic changes with expression of 355 and 294 genes differentially regulated by serotype 6B (BHN418) and serotype 1 (BVJ1JL) respectively (**Figure 6.13**) after 10 minutes. In contrast to iron-deplete growth conditions, growth of both serotype 6B and serotype 1 in CDM supplemented with iron (1mM FeSO<sub>4</sub>) downregulated the expression of *piu* (*piuABCD*) operon while the *cop* (*copAY*, *cupA*) operon was not differentially expressed, supporting the involvement of these loci in *S. pneumoniae* metal ion homeostasis as previously described (Brown, Gilliland, and Holden 2001; Shafeeq *et al.* 2011). The expression of the iron storage protein *dpr* was also reversed and upregulated in iron replete CDM compared to iron deplete conditions (**Table 6.10A** and **Table 6.10B**). This is consistent with the homeostatic function of *dpr* in *S. pneumoniae* iron regulation. The expression of the branched-chain amino acid transport system 2 carrier protein, *brnQ*, was also reversed, being downregulated, or not differentially expressed in iron supplemented CDM. The *gls24* gene (*SPD\_1509* in D39 strain and *JN057\_RS08665* in BVJ1JL strain), recently described as a novel pneumococcal iron transporter, and suggested to function similar to *piuABCD* was not downregulated in iron-replete conditions (**Table 6.10A**) (Miao *et al.* 2018). This is in contrast to *piuABCD* operon the regulation of which was correlated with iron-deplete/iron-replete growth conditions in this work. The *fucK* gene, which is involved in the phosphorylation and degradation of L-fucose was the most significantly upregulated gene by serotype 6B in early response to iron-repletion (**Table 6.10A**). The *fucK* gene was also significantly upregulated by serotype 1 in iron replete media (**Table 6.10B**). In iron depleted THY, the *fucK* gene was not differentially expressed by both strains. Despite the contrasting growth environments of iron-deplete and iron-replete conditions, several genes were upregulated in both conditions (**Table 6.9A**, **Table 6.9B**, **Table 6.10A**, **Table 6.10B**), and these mainly included the competence sensing genes (*ciaR*, *ciaH*, *coiA*, *comA*), pyrimidine biosynthesis genes (*carA*, *carB*), several ribosomal proteins (e.g., *rplC*, *rlmN*, *rplV*, and *rplB*), carbohydrate metabolism genes (e.g., *malQ*), and the putative iron response genes *JN057\_RS03040*, *JN057\_RS04360*, and *JN057\_RS04930* (**Table 6.10A** and **Table 6.10B**).

## DGE at 10 mins (CDM): Iron-replete Conditions



**Figure 6.13: Differential gene expression (DGE) of BHN418 (serotype 6B) and BVJ11JL (serotype 1) after 10 minutes in iron replete CDM.** Volcano (A – B) and bar (C –D) plots showing up- (pink) and down-regulated (purple) genes in serotype 6B and serotype 1 strains. Each point in the volcano plots indicate gene expression with colour indicative of expression direction (pink = upregulated, purple = downregulated). Where available in the original file, gene names were annotated on volcano plots, allowing maximum of 40 gene annotations on each plot. **(E)** Venn diagram showing overlap in number of differentially expressed genes (DEGs) between serotype 6B and serotype 1 at this condition. DEGs were defined by fold change  $>1.5$  and false discovery rate  $\text{FDR} < 0.05$ .

**Table 6.10A:** Top 35 differentially upregulated genes by *S. pneumoniae* serotype 6B (BHN418) after 10 minutes growth in iron replete CDM (1mM FeSO<sub>4</sub>). Genes were ranked by statistical significance (P-adj). Bold gene names were upregulated in both serotype 1 and serotype 6B. RefSeq locus tags of published sequenced of BVJ1JL are shown.

locus	Gene	FC	P-adj	Annotation
<b>Serotype 6B (BHN418)</b>				
<i>JN057_RS10680</i>	<i>fucK</i>	10.98	0.00E+00	L-fuculokinase
<b><i>JN057_RS03040</i></b>		4.31	0.00E+00	M57 family metalloprotease
<i>JN057_RS07620</i>	<i>dpr</i>	4.28	0.00E+00	DNA starvation/stationary phase protection protein/ Bifunctional non-specific DNA-binding protein/ferroxidase Dpr
<b><i>JN057_RS03870</i></b>	<b><i>ciaR</i></b>	3.74	0.00E+00	two-component system response regulator CiaR
<b><i>JN057_RS04360</i></b>		4.41	4.84E-273	PspC domain-containing protein
<b><i>JN057_RS03875</i></b>	<b><i>ciaH</i></b>	2.86	7.78E-253	two-component system sensor histidine kinase CiaH
<b><i>JN057_RS04930</i></b>		2.34	5.23E-201	DUF1002 domain-containing protein
<b><i>JN057_RS11045</i></b>	<b><i>htrA</i></b>	2.48	3.84E-185	Serine protease, DegP/HtrA
<b><i>JN057_RS01185</i></b>	<b><i>folC</i></b>	2.99	5.65E-180	bifunctional folylpolyglutamate synthase/dihydrofolate synthase
<i>JN057_RS10090</i>		3.34	3.76E-141	DUF1304 domain-containing protein
<i>JN057_RS10395</i>	<i>malQ</i>	2.67	1.97E-131	4-alpha-glucanotransferase
<i>JN057_RS08475</i>		3.19	2.41E-124	DUF4649 family protein
<i>JN057_RS08655</i>	<i>amaP</i>	2.52	6.30E-124	alkaline shock response membrane anchor protein AmaP
<i>JN057_RS02550</i>	<i>glnA</i>	2.33	5.41E-106	type I glutamate--ammonia ligase
<i>JN057_RS01340</i>	<i>adK</i>	2.54	2.04E-105	adenylate kinase
<i>JN057_RS10755</i>	<i>glpO</i>	2.50	7.75E-102	type 1 glycerol-3-phosphate oxidase
<i>JN057_RS10760</i>	<i>glpK</i>	2.75	1.84E-101	glycerol kinase GlpK
<i>JN057_RS01175</i>	<i>ruvX</i>	1.84	3.37E-100	Holliday junction resolvase RuvX
<i>pcp_2</i>	<i>pcp</i>	2.18	1.38E-99	(Table Continued)

<i>JN057_RS08665</i>	<i>gls24</i>	2.10	5.41E-97	Asp23/Gls24 family envelope stress response protein
<b><i>JN057_RS04375</i></b>		2.11	2.65E-96	ABC transporter permease
<i>JN057_RS07130</i>	<i>pdxS</i>	2.15	1.19E-92	pyridoxal 5'-phosphate synthase lyase subunit PdxS
<i>JN057_RS08650</i>		2.89	2.06E-92	GlsB/YeaQ/YmgE family stress response membrane protein
<b><i>JN057_RS10750</i></b>	<b><i>glpF</i></b>	2.39	2.59E-89	Glycerol uptake facilitator protein
<i>JN057_RS04535</i>	<i>cvfB</i>	2.07	3.89E-89	RNA-binding virulence regulatory protein CvfB
<i>JN057_RS10095</i>		3.48	2.53E-85	MarR family winged helix-turn-helix transcriptional regulator
<i>JN057_RS08660</i>		2.76	1.20E-78	Small integral membrane protein
<b><i>JN057_RS02085</i></b>	<b><i>liaF</i></b>	2.02	3.67E-76	cell wall-active antibiotics response protein
<i>JN057_RS01190</i>		2.05	1.21E-75	SP_0198 family lipoprotein
<i>JN057_RS08290</i>	<i>qsrB</i>	1.94	1.52E-71	ABC transporter membrane-spanning permease – Na <sup>+</sup> export
<i>JN057_RS08295</i>	<i>qsrA</i>	1.89	8.76E-71	ABC transporter ATP-binding protein- Na <sup>+</sup> export
<i>JN057_RS04660</i>	<i>pmrA</i>	1.91	5.79E-70	multidrug efflux MFS transporter PmrA
<b><i>JN057_RS02545</i></b>	<b><i>glnR</i></b>	2.45	9.53E-70	transcriptional repressor GlnR
<i>JN057_RS04595</i>	<i>infC</i>	2.19	1.55E-69	translation initiation factor IF-3
<b><i>JN057_RS11050</i></b>	<b><i>parB</i></b>	1.93	9.35E-67	ParB/RepB/Spo0J chromosome partition protein

**Table 6.10B:** Top 35 differentially upregulated genes by *S. pneumoniae* serotype 1 (BVJ1JL) after 10 minutes incubation in iron replete CDM (1mM FeSO<sub>4</sub>). Genes were ranked by statistical significance (P-adj). Bold gene names are upregulated in both serotype 1 and serotype 6B. RefSeq locus tags of BVJ1JL shown.

locus	Gene	FC	P-adj	Annotation
<b>Serotype 1 (BVJ1JL)</b>				
<b>JN057_RS04930</b>		4.06	0.00E+00	DUF1002 domain-containing protein
<b>JN057_RS03870</b>	<b><i>ciaR</i></b>	3.91	0.00E+00	two-component system response regulator CiaR
<b>JN057_RS03040</b>		3.91	9.63E-277	M57 family metalloprotease
<b>JN057_RS03875</b>	<b><i>ciaH</i></b>	3.34	1.95E-276	two-component system sensor histidine kinase CiaH
<i>JN057_RS10390</i>	<i>glgP</i>	2.76	2.50E-231	glycogen/starch/alpha-glucan family phosphorylase
<i>JN057_RS03005</i>	<i>glnH2</i>	3.92	1.90E-220	Glutamine transporter substrate- binding protein
<i>JN057_RS10395</i>	<i>malQ</i>	3.58	2.82E-213	4-alpha-glucanotransferase
<b>JN057_RS11045</b>	<b><i>htrA</i></b>	3.69	2.30E-148	Serine protease, DegP/HtrA
<i>JN057_RS10680</i>	<i>fucK</i>	8.23	2.62E-146	L-fuculokinase
<i>JN057_RS02690</i>	<i>blpA</i>	4.07	5.12E-146	peptide cleavage/export ABC transporter BlpA
<i>JN057_RS10675</i>		6.77	5.61E-143	ABC transporter substrate-binding protein
<i>JN057_RS03000</i>	<i>glnP2a</i>	3.88	1.50E-137	Glutamine amino acid ABC transporter permease
<b>JN057_RS04360</b>		3.26	1.11E-123	PspC domain-containing protein
<i>JN057_RS09585</i>		2.03	9.63E-120	Bax inhibitor-1/YccA family protein
<i>JN057_RS03010</i>		4.33	3.19E-112	amino acid ABC transporter ATP-binding protein
<i>JN057_RS07130</i>	<i>pdxS</i>	1.96	2.11E-111	pyridoxal 5'-phosphate synthase lyase subunit PdxS

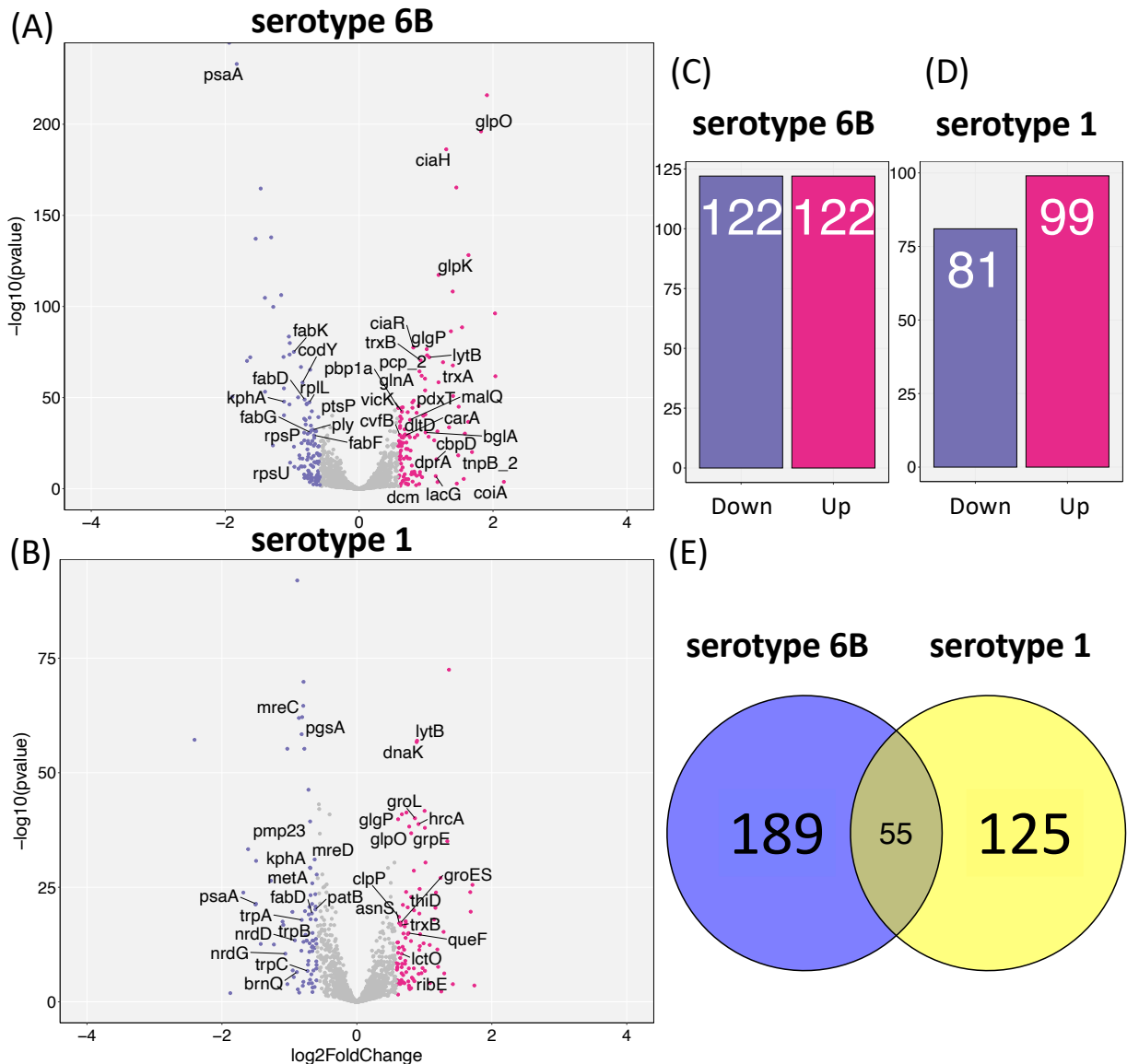


<i>(Table Continued)</i>				
<i>JN057_RS02550</i>	<i>glnA</i>	2.87	6.08E-108	type I glutamate--ammonia ligase
<i>JN057_RS02995</i>	<i>glnP2b</i>	3.68	5.34E-101	Glutamine ABC transporter permease
<i>JN057_RS02695</i>	<i>blpM</i>	3.23	1.48E-100	two-peptide bacteriocin subunit BlpM
<i>JN057_RS10755</i>	<i>glpO</i>	2.29	2.42E-91	type 1 glycerol-3-phosphate oxidase
<i>JN057_RS07680</i>		2.01	4.21E-89	DEAD/DEAH box helicase
<i>JN057_RS10725</i>	<i>dltA</i>	1.64	2.26E-83	D-alanine--poly(phosphoribitol) ligase subunit DltA
<i>JN057_RS02685</i>	<i>blpB</i>	3.29	1.92E-77	bacteriocin secretion accessory protein
<i>JN057_RS02545</i>	<i>glnR</i>	2.76	5.77E-76	transcriptional repressor GlnR
<i>JN057_RS11050</i>	<i>parB</i>	2.46	1.15E-75	ParB/RepB/Spo0J family partition protein
<i>JN057_RS06325</i>		1.70	2.46E-70	alcohol dehydrogenase catalytic domain-containing protein
<i>JN057_RS07125</i>	<i>pdxT</i>	1.96	2.83E-69	pyridoxal 5'-phosphate synthase glutaminase subunit PdxT
<i>JN057_RS01160</i>	<i>spx</i>	1.85	3.21E-69	transcriptional regulator Spx
<i>JN057_RS02085</i>	<i>liaF</i>	1.91	3.75E-69	cell wall-active antibiotics response protein
<i>JN057_RS06190</i>	<i>glnQ</i>	2.79	8.20E-64	Glutamine amino acid ABC transporter ATP-binding protein
<i>JN057_RS10750</i>	<i>gnpF</i>	2.35	2.69E-63	Glycerol uptake facilitator protein
<i>JN057_RS02095</i>	<i>vraR</i>	1.87	4.14E-63	Two-component transcriptional regulator vraR
<i>JN057_RS04375</i>		2.37	9.14E-61	ABC transporter permease
<i>JN057_RS04230</i>		2.63	4.84E-60	hypothetical protein
<i>blpC_2</i>	<i>blpC</i>	3.54	6.64E-56	

### 6.3.14 Later transcriptomic response of *S. pneumoniae* in iron replete media

After 50 minutes incubation in iron replete CDM, 122 genes were upregulated by the serotype 6B strain while 99 genes were upregulated by serotype 1 (**Figure 6.14**). The expression of pneumococcal iron storage protein *dpr* remained significantly upregulated after 50 minutes by both strains (**Table 6.11A** and **Table 6.11B**). A high-Fe environment can cause increased oxidative stress to bacteria (Ong *et al.* 2013). The lactate oxidase gene (*lctO*) and pyruvate oxidase (*spxB*) are the main bacterial enzymes involved in pneumococcal resistance to oxidative stress and implicated as virulence factors for *S. pneumoniae* colonisation (Mraheil *et al.* 2021). *lctO* expression was significantly upregulated by serotype 1 (FC = 1.60, FDR = 3.25E-15) but not in serotype 6B (FC = 1.01, FDR = 0.95) at 50 mins incubation in iron replete conditions. *spxB* on the other hand was not differentially expressed by either strain at this timepoint or in the earlier iron replete-timepoints. However, *psaA*, which also plays a role in oxidative resistance was downregulated by both strains at 50 minutes in iron rich media (Tseng *et al.* 2002). *psaA* is a manganese import system and a virulence gene for pneumococcal colonisation (Tseng *et al.* 2002). In iron deplete media, *psaA* expression was upregulated by both serotype 1 and serotype 6B indicating sensitivity of its gene regulation to iron availability. Ranking of the top 35 genes upregulated by serotype 6B strain and serotype 1 at 50 minutes in iron replete conditions are listed in **Table 6.11A** and **Table 6.11B** respectively.

## DGE at 50 mins (CDM): Iron-replete Conditions



**Figure 6.14: Differential gene expression (DGE) of BHN418 (serotype 6B) and BVJ1JL (serotype 1) after 50 minutes in iron replete CDM.** Volcano (A – B) and bar plots (C – D) showing up- (pink) and down-regulated (purple) genes in serotype 6B and serotype 1 strains. Each point in the volcano plots indicate gene expression with colour indicative of expression direction (pink = upregulated, purple = downregulated). Where available in the original file, gene names are annotated on volcano plots, allowing maximum of 40 gene annotations on each plot. **(E)** Venn diagram showing overlap in number of DEGs between serotype 6B and serotype 1 at this condition. DEGs were defined by fold change  $>1.5$  and false discovery rate FDR  $<0.05$ .

**Table 6.11A:** Top 35 differentially upregulated genes by *S. pneumoniae* serotype 6B (BHN418) after 50 minutes incubation in iron replete CDM (1mM FeSO<sub>4</sub>). Genes were ranked by statistical significance (P-adj). Bold gene names are upregulated in both serotype 1 and serotype 6B. RefSeq locus tags of BVJ1JL are shown.

locus	Gene	FC	P-adj	Annotation
<b>Serotype 6B (BHN418)</b>				
<i>JN057_RS10750</i>	<i>glpF</i>	3.76	1.16E-216	Glycerol uptake facilitator protein
<i>JN057_RS10755</i>	<i>glpO</i>	3.53	8.09E-197	type 1 glycerol-3-phosphate oxidase
<i>JN057_RS03875</i>	<i>ciaH</i>	2.47	5.82E-187	two-component system sensor histidine kinase CiaH
<i>JN057_RS07620</i>	<i>dpr</i>	2.74	5.23E-166	DNA starvation/stationary phase protection protein/ Bifunctional non-specific DNA-binding protein/ferroxidase Dpr
<i>JN057_RS10760</i>	<i>glpK</i>	3.10	6.62E-129	glycerol kinase GlpK
<b><i>JN057_RS04375</i></b>		2.28	5.19E-118	ABC transporter permease
<i>JN057_RS07495</i>		2.64	6.17E-109	NAD(P)H-dependent oxidoreductase
<i>JN057_RS10635</i>	<i>fucl</i>	4.09	6.56E-97	Bifunctional L-fucose isomerase/D-arabinose ketol-isomerase
<i>JN057_RS10680</i>	<i>fucK</i>	2.90	2.79E-89	L-fuculokinase
<i>JN057_RS10090</i>		2.59	4.17E-87	DUF1304 domain-containing protein
<i>JN057_RS03870</i>	<i>ciaR</i>	1.75	3.51E-78	two-component system response regulator CiaR
<i>JN057_RS10390</i>	<i>glgP</i>	2.01	2.72E-77	glycogen/starch/alpha-glucan family phosphorylase
<i>JN057_RS01185</i>	<i>folC</i>	2.02	6.22E-74	bifunctional folylpolyglutamate synthase/dihydrofolate synthase
<b><i>JN057_RS04625</i></b>	<b><i>lytB</i></b>	2.07	7.52E-73	glucosaminidase domain-containing protein
<i>JN057_RS07080</i>	<i>trxB</i>	1.89	3.87E-71	thioredoxin-disulfide reductase
<i>JN057_RS08475</i>		2.38	4.08E-70	DUF4649 family protein
<i>JN057_RS03000</i>	<i>glnP2a</i>	2.64	2.28E-68	Glutamine amino acid ABC transporter permease

(Table Continued)

<i>pcp_2</i>		1.87	4.43E-65	
<i>JN057_RS02550</i>	<i>glnA</i>	1.91	1.07E-62	type I glutamate--ammonia ligase
<i>group_1261</i>		4.10	1.94E-62	hypothetical protein
<i>JN057_RS08970</i>	<i>proWX</i>	1.98	4.15E-61	ABC transporter permease/substrate-binding protein
<i>JN057_RS08480</i>	<i>trxA</i>	2.28	4.04E-59	thioredoxin
<i>JN057_RS07125</i>	<i>pdxT</i>	1.98	1.14E-54	pyridoxal 5'-phosphate synthase glutaminase subunit PdxT
<i>JN057_RS10095</i>		2.65	1.38E-51	MarR family winged helix-turn-helix transcriptional regulator
<i>JN057_RS00785</i>		1.76	4.26E-49	HAD family hydrolase
<i>JN057_RS03955</i>		1.74	4.31E-48	ATP-dependent Clp protease ATP-binding subunit
<i>JN057_RS08975</i>	<i>proV</i>	1.85	3.92E-46	Glycine betaine/L-proline ABC transporter ATP-binding protein
<b><i>JN057_RS00625</i></b>		2.80	1.03E-45	Glyoxalase family protein
<i>JN057_RS07105</i>		1.57	2.28E-45	methylated-DNA--[protein]-cysteine S- methyltransferase
<i>JN057_RS07690</i>		1.55	3.64E-45	FAD-containing oxidoreductase
<i>JN057_RS01190</i>		1.74	4.00E-45	SP_0198 family lipoprotein
<i>JN057_RS01990</i>	<i>pbp1a</i>	1.50	5.56E-45	penicillin-binding protein PBP1A
<i>JN057_RS06115</i>	<i>vicK</i>	1.56	6.18E-43	cell wall metabolism sensor histidine kinase Vick
<i>JN057_RS04340</i>		1.65	1.05E-42	permease
<i>JN057_RS06110</i>	<i>vicX</i>	1.81	2.42E-42	Zn-dependent hydrolase

**Table 6.11B:** Top 35 differentially upregulated genes by *S. pneumoniae* serotype 1 (BVJ1JL) after 50 minutes incubation in iron replete CDM (1mM FeSO<sub>4</sub>). Genes were ordered by statistical significance (P-adj). Bold gene names are upregulated in both serotype 1 and serotype 6B. RefSeq locus tags of BVJ1JL are shown.

locus	Gene	FC	P-adj	Annotation
<b>Serotype 1 (BVJ1JL)</b>				
<b>JN057_RS04375</b>		2.58	3.31E-73	ABC transporter permease
<b>JN057_RS04625</b>	<b>lytB</b>	1.86	1.03E-57	glucosaminidase domain-containing protein
<i>dnaK</i>	<i>dnaK</i>	1.85	2.04E-57	molecular chaperone DnaK
<i>JN057_RS10750</i>		2.01	2.12E-42	aquaporin family protein
<i>JN057_RS08075</i>		1.66	4.84E-42	MurR/RpiR family transcriptional regulator
<i>JN057_RS04930</i>		1.59	1.15E-41	DUF1002 domain-containing protein
<i>JN057_RS09315</i>	<i>groL</i>	1.82	8.70E-41	chaperonin GroEL
<i>JN057_RS10390</i>	<i>glgP</i>	1.53	1.43E-40	glycogen/starch/alpha-glucan family phosphorylase
<i>JN057_RS02610</i>	<i>hrcA</i>	1.88	1.61E-39	heat-inducible transcriptional repressor HrcA
<i>JN057_RS10755</i>	<i>glpO</i>	1.71	5.68E-39	type 1 glycerol-3-phosphate oxidase
<i>JN057_RS02615</i>	<i>grpE</i>	2.01	1.06E-38	nucleotide exchange factor GrpE
<i>JN057_RS02990</i>		1.75	1.62E-37	ferric reductase-like transmembrane domain-containing protein
<b>JN057_RS04370</b>		2.53	8.76E-36	ABC transporter ATP-binding protein
<i>JN057_RS10635</i>	<i>fucl</i>	2.02	3.99E-31	Bifunctional L-fucose isomerase/D-arabinose ketol-isomerase
<i>JN057_RS00800</i>		1.80	2.41E-29	LysM peptidoglycan-binding domain-containing protein
<i>JN057_RS10675</i>		2.36	9.19E-28	ABC transporter substrate-binding protein
<i>group_1142</i>		3.28	2.97E-26	unannotated protein
<i>JN057_RS02995</i>	<i>glnP2b</i>	1.91	2.35E-25	amino acid ABC transporter permease

(Table Continued)

<i>JN057_RS05655</i>		1.66	1.05E-24	bifunctional riboflavin kinase/FAD synthetase
<b><i>JN057_RS00625</i></b>		3.21	1.15E-24	Glyoxalase family protein
<i>JN057_RS07495</i>		2.25	1.38E-24	NAD(P)H-dependent oxidoreductase
<i>JN057_RS07765</i>		1.75	1.36E-23	ferredoxin
<i>JN057_RS03880</i>		1.60	7.03E-22	DUF3270 domain-containing protein
<i>JN057_RS08890</i>		1.69	2.44E-21	hypothetical protein
<i>JN057_RS10680</i>	<i>fucK</i>	2.24	3.08E-21	L-fuculokinase
<i>JN057_RS03535</i>	<i>thiD</i>	1.80	1.08E-20	bifunctional hydroxymethylpyrimidine kinase/phosphomethylpyrimidine kinase
<i>JN057_RS07620</i>	<i>dpr</i>	3.22	2.15E-20	DNA starvation/stationary phase protection protein/ Bifunctional non-specific DNA-binding protein/ferroxidase Dpr
<i>JN057_RS10100</i>		1.90	5.46E-20	LysM peptidoglycan-binding domain-containing protein
<i>JN057_RS03620</i>	<i>clpP</i>	1.54	2.98E-19	ATP-dependent Clp protease proteolytic subunit ClpP
<i>JN057_RS10650</i>		2.21	1.07E-18	alpha-galactosidase
<i>JN057_RS01070</i>		1.65	2.65E-18	hypothetical protein
<i>JN057_RS09320</i>	<i>groES</i>	1.57	5.18E-18	co-chaperone GroES
<i>JN057_RS07475</i>	<i>asnS</i>	1.55	7.05E-18	asparagine--tRNA ligase
<i>JN057_RS01840</i>		1.66	7.77E-18	ATP-dependent Clp protease ATP-binding subunit
<i>JN057_RS07080</i>	<i>trxB</i>	1.57	1.81E-17	thioredoxin-disulfide reductase

### 6.3.15 Orthologous genes in *S. pneumoniae* strains

Given the large number of differentially expressed genes (DEGs) in CDM experiments, we investigated the biological pathways that were enriched by BHN418 and BVJ1JL using *S. pneumoniae* KEGG pathways database. For BHN418 we used the annotation of serotype 6B strain 670-6B and for BVJ1JL we used the annotation for serotype 1 strain INV104. **Table 6.12** shows the proportion of DEGs in our study strains that were present/annotated in reference strains and used for pathway analysis.

**Table 6.12:** The proportion of differentially expressed genes of BHN418UCL (serotype 6B) and BVJ1JL (serotype 1) that were present/annotated in KEGG reference genomes used for pathway enrichment analyses.

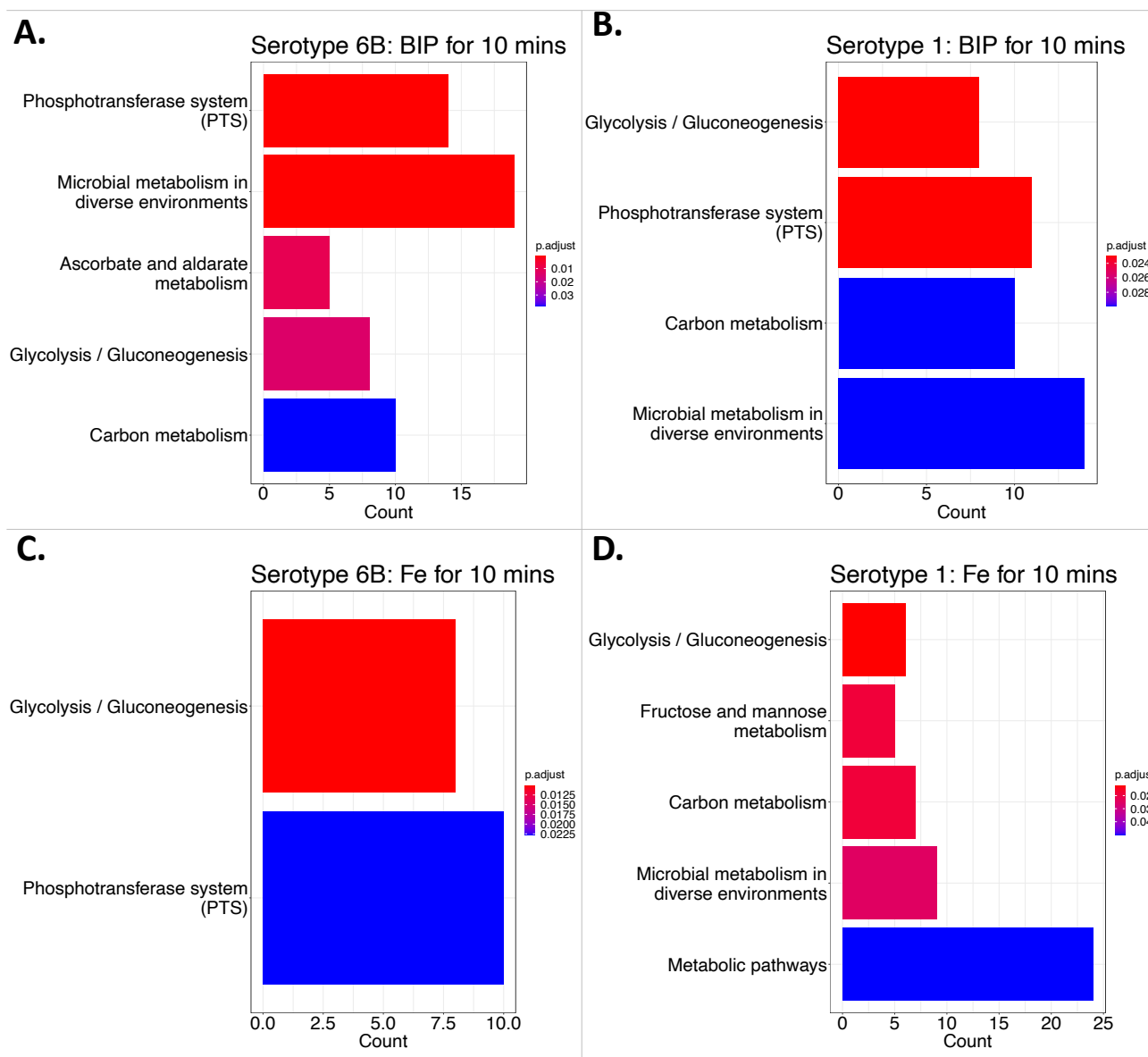
Condition	Media	Serotype 6B		Serotype 1	
		BHN418UCL	670-6B	BVJ1JL	INV104
Upregulated genes					
BIP10	CDM	196	172 (87.8%)	176	146 (82.9%)
BIP50	CDM	104	90 (86.5%)	101	72 (71.3%)
FeSO10	CDM	187	165 (88.2%)	170	133 (78.2%)
FeSO50	CDM	122	109 (89.3%)	99	72 (72.7%)
BIP10	THY	17	17 (100.0%)	79	67 (84.8%)
BIP50	THY	73	62 (84.9%)	41	35 (85.4%)
Downregulated genes					
BIP10	CDM	181	147 (81.2%)	190	144 (75.8%)
BIP50	CDM	103	92 (89.3%)	72	51 (70.8%)
FeSO10	CDM	168	138 (82.1)	124	95 (76.6%)
FeSO50	CDM	122	104 (85.2%)	81	59 (72.8%)
BIP10	THY	28	28 (100%)	102	92 (90.2%)
BIP50	THY	51	46 (90.2%)	36	29 (80.6%)



### 6.3.16 Regulated pathways by *S. pneumoniae* during early adaptation to iron-deplete and iron-replete conditions in CDM

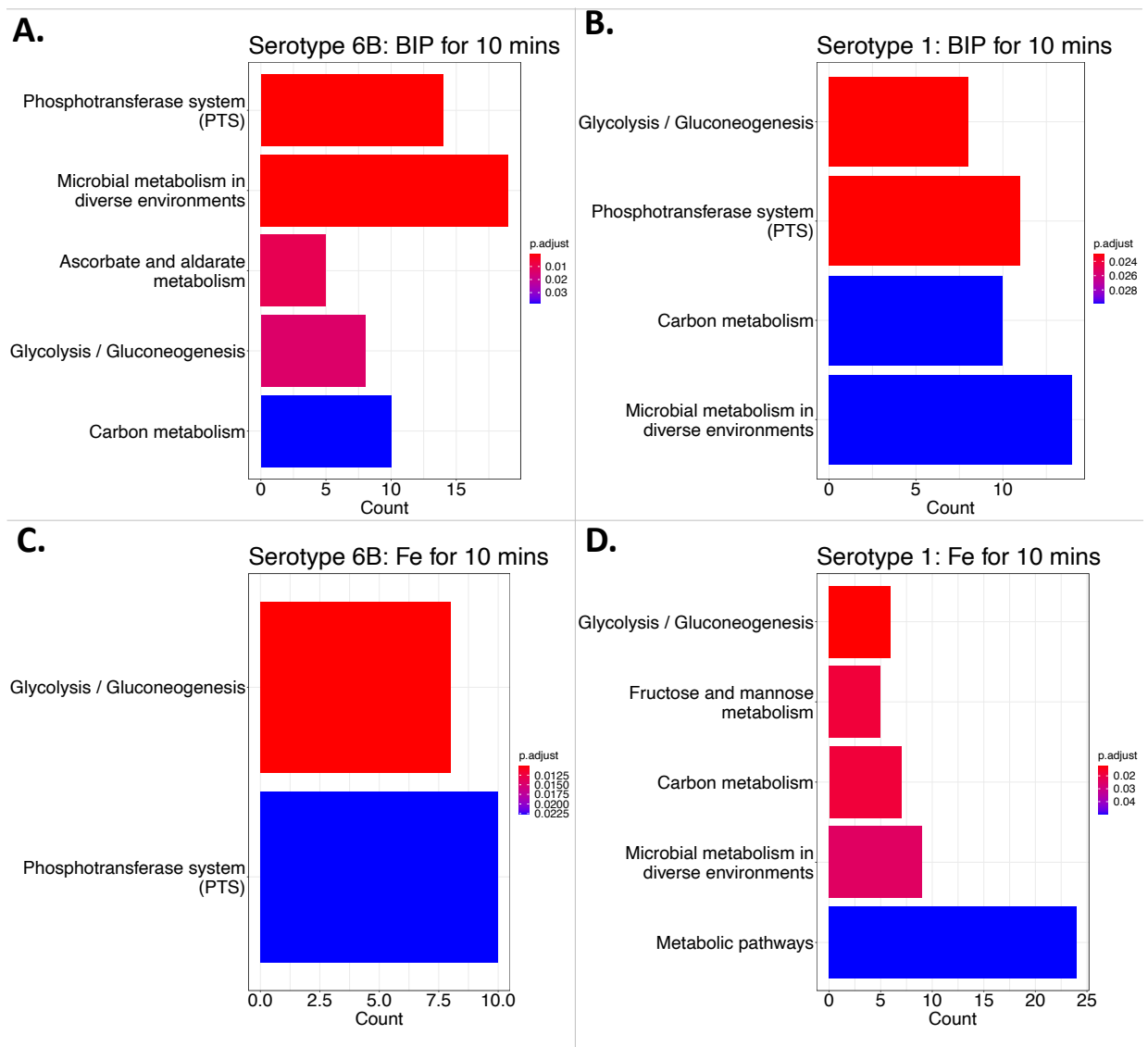
Pneumococcal two-component system, ribosome synthesis, and biosynthesis of secondary cofactors were the most differentially regulated pathways by both serotype 1 and serotype 6B in iron-deplete and iron-replete conditions at 10 minutes (**Figure 6.15**). Strain-specific pathways that were enriched included the oxidative phosphorylation pathway which was upregulated in serotype 6B but not in serotype 1 in iron-deplete conditions (**Figure 6.15**). In iron-replete media, serotype 1 uniquely unregulated pathways involved in pyrimidine metabolism during early adaptation (**Figure 6.15**). The pathways that were significantly enriched by downregulated genes in iron-deplete and iron-replete conditions by both strains are shown in **Figure 6.16**. Pneumococcal phosphotransferase system (PTS), carbon metabolism, and gluconeogenesis biological pathways were differentially downregulated by both strains in iron-deplete and iron-replete conditions (**Figure 6.16**).

## Significantly Upregulated Pathways : CDM 10 mins



**Figure 6.15: Upregulated biological pathways by *S. pneumoniae* serotype 6B (BHN418) and serotype 1 (BVJ1JL) in iron-deplete (A and B) and iron-replete (C and D) CDM at 10 minutes.** Pathway enrichment was performed by over representation analysis. Pathways with adjusted p-value <0.05 were significantly enriched. All pathways shown in the plot are significantly enriched with the length of the bar corresponding to the number of DEGs in respective pathways. The colour of the bars denotes level of statistical significance based on adjusted p-value with red being most significantly enriched pathways.

## Significantly Downregulated Pathways : CDM 10 mins



**Figure 6.16: Downregulated biological pathways by *S. pneumoniae* serotype 6B (BHN418) and serotype 1 (BVJ1JL) in iron-deplete (A and B) and iron-replete (C and D) conditions at 10 minutes.** Pathway enrichment was performed by over representation analysis. Pathways with adjusted p-value <0.05 were significantly enriched. All pathways shown in the plot are significantly enriched with the length of the bar corresponding to the number of DEGs in respective pathways. The colour of the bars denotes level of statistical significance based on adjusted p-value with red being most significantly enriched pathways.

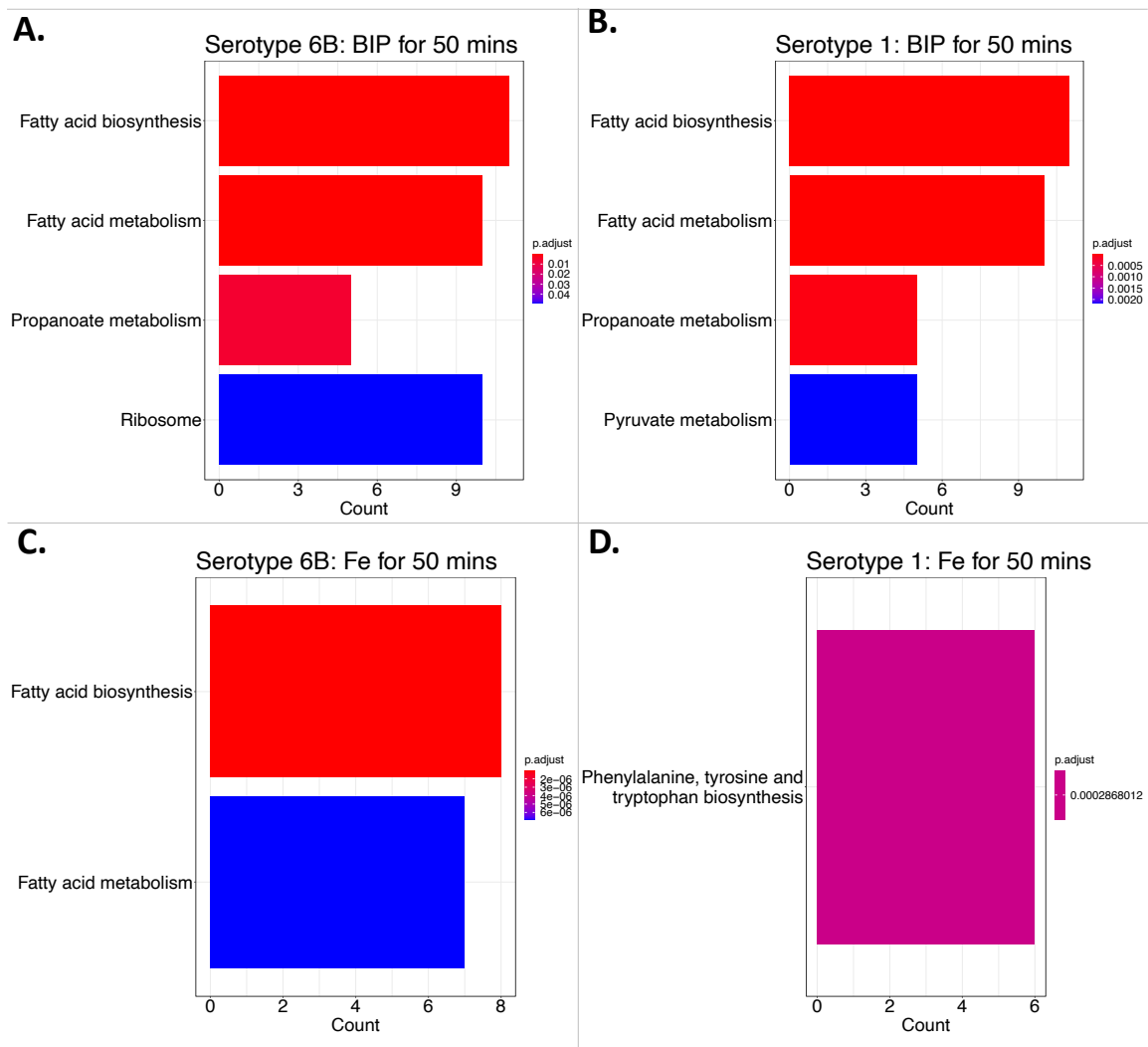
### 6.3.17 Regulated pathways by *S. pneumoniae* during late adaptation to iron-deplete and iron-replete conditions in CDM

At 50 minutes incubation in iron-deplete CDM, serotype 6B upregulated two-component system pathways while serotype 1 upregulated purine metabolism and secondary metabolites biosynthesis pathways (**Table 6.13**). Two-component systems pathways were also upregulated by serotype 6B during growth in iron-replete CDM at 50 minutes (**Table 6.13**). No biological pathways were enriched by the upregulated genes of serotype 1 in iron replete CDM at 50 minutes. Continued growth in iron-deplete CDM up to 50 minutes significantly downregulated fatty acid biosynthesis and propanoate metabolism pathways in both serotype 1 and serotype 6B (**Figure 6.17**). Fatty acid-biosynthesis pathways were also significantly downregulated by serotype 6B at 50 minutes in iron-replete conditions but not by serotype 1. Serotype 1 uniquely downregulated *S. pneumoniae* pathways involved with phenylalanine, tyrosine, and tryptophan biosynthesis at 50 minutes growth in iron limited CDM (**Figure 6.17**).

**Table 6.13** Biological pathways differentially upregulated by *S. pneumoniae* serotype 1 (BVJ1JL) and serotype 6B (BHN418) strains after 50 minutes growth in iron-deplete and iron-replete conditions in CDM. Orthologs of KEGG reference strains INV104 and 670-6B were used where available.

Pathway ID	Description	GeneRatio	BgRatio	p.adjust	qvalue
<b>Upregulated pathways by Serotype 1 (BVJ1JL), 50 mins in iron-deplete CDM</b>					
sni01110	Biosynthesis of secondary metabolites	17/30	153/732	2.40E-04	2.15E-04
sni00230	Purine metabolism	8/30	35/732	3.31E-04	2.96E-04
<b>Upregulated pathways by Serotype 6B (BHN418), 50 mins in iron-deplete CDM</b>					
snb02020	Two-component system	10/38	40/746	2.28E-04	2.19E-04
<b>Upregulated pathways by Serotype 6B (BHN418), 50 mins in iron replete CDM</b>					
snb02020	Two-component system	11/54	40/746	1.27E-03	1.19E-03

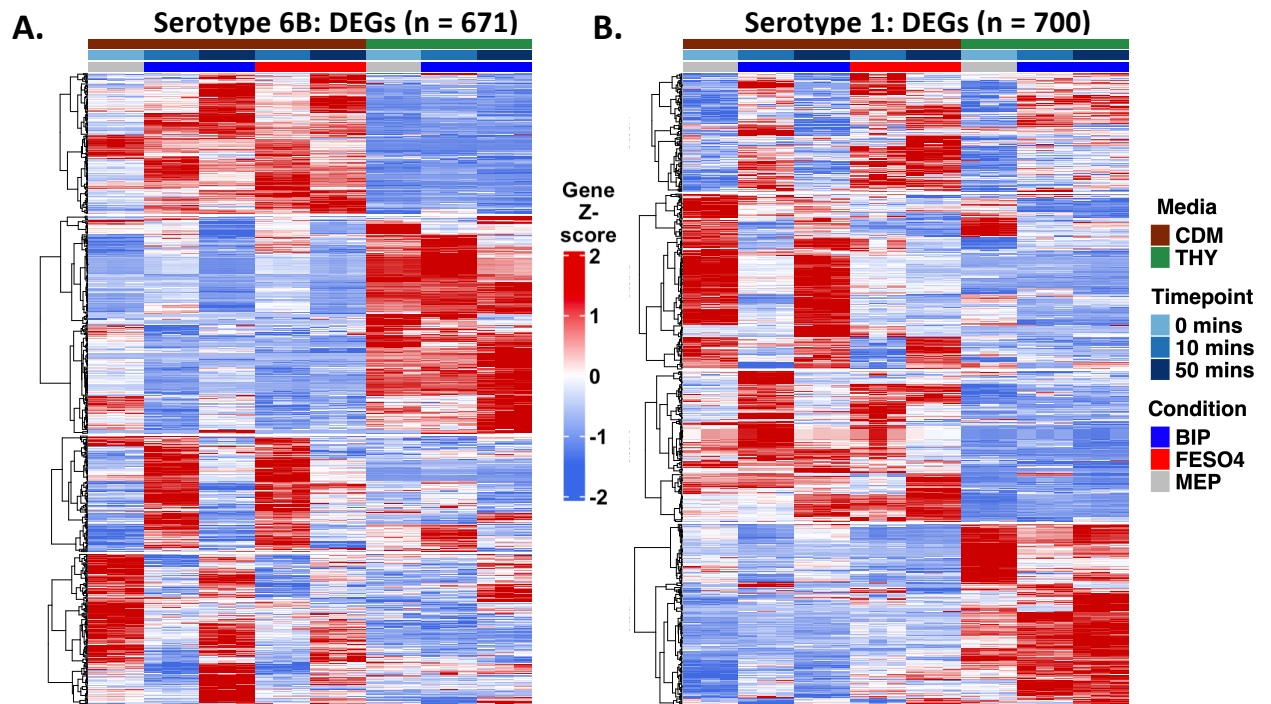
## Significantly Downregulated Pathways : CDM 50 mins



**Figure 6.17: Downregulated biological pathways by *S. pneumoniae* serotype 6B (BHN418) and serotype 1 (BVJ1JL) in iron-deplete (A and B) and iron-replete (C and D) conditions at 50 minutes.** Pathway enrichment was performed by over representation analysis. Pathways with adjusted p-value <0.05 were significantly enriched. All pathways shown in the plot are significantly enriched with the length of the bar corresponding to the number of DEGs in respective pathways. The colour of the bars denotes level of statistical significance based on adjusted p-value with red being most significantly enriched pathways.

### 6.3.18 Strain specific transcriptomics changes during growth in iron-deplete and iron-replete conditions

To understand gene expression changes across all conditions for each strain, unsupervised hierarchical heatmap clustering of genes that were differentially expressed in at least one experimental condition was performed separately for each strain. In total 671 and 700 genes were differentially expressed in at least one condition for serotype 6B (BHN418) and serotype 1 (BVJ1JL) strains respectively. For each of the strains, four main gene clusters were identified highlighting the dynamic expression pattern of *S. pneumoniae* in iron deplete-and iron-replete conditions (**Figure 6.18**). While some genes reverse their gene expression in response to iron availability, some genes were upregulated in both conditions (relative to the control samples) (**Figure 6.18**). **Table 6.14** shows an example of selected genes that showed differential expression levels between BHN418 (serotype 6B) and BVJ1JL (serotype 1) in iron-deplete CDM at 10 minutes.



**Figure 6.18: Unsupervised hierarchical clustering analysis of gene expression of *S. pneumoniae* serotype 6B (BHN418UCL) and serotype 1 (BVJ1JL) in iron-deplete and iron-replete conditions.** Heatmap based on z-scores of gene expression was constructed for all genes that were differentially expressed in at least one study condition (n = 671 for serotype 6B and n = 700 for serotype 1). Upper tracks and legends indicate triplicates samples of each condition.

**Table 6.14:** Selected *S. pneumoniae* genes showing different expression levels between *S. pneumoniae* serotype 6B (BHN418) and serotype 1 (BVJ1JL) strains at 10 minutes growth in iron-deplete CDM.

Gene	Serotype 1 FC	Serotype 6B FC	Annotation	D39V Ortholog
<i>blpM</i>	3.09	-1.10	two-peptide bacteriocin subunit BlpM	No
<i>blpN</i>	3.32	-1.18	two-peptide bacteriocin subunit BlpN	No
<i>cbpD</i>	1.59	2.84	choline binding-anchored murein hydrolase CbpD	Yes
<i>coiA</i>	1.35	2.57	competence protein CoiA	Yes
<i>JN057_RS01415</i>	-1.35	-2.06	glycyl-radical enzyme activating protein	Yes
<i>JN057_RS01750</i>	-1.57	-2.10	RpiB/LacA/LacB family sugar-phosphate isomerase	Yes
<i>JN057_RS01755</i>	-1.36	-2.09	gluconate 5-dehydrogenase	Yes
<i>blpT</i>	3.68	-1.11	hypothetical protein	Yes
<i>blpB</i>	3.54	-1.44	bacteriocin secretion accessory protein	No
<i>blpL</i>	2.71	1.30	Immunity protein BlpL	No
<i>blpX</i>	2.09	1.18	Immunity protein BlpX	No
<i>blpY</i>	2.04	1.22	CPBP family intramembrane metalloprotease	Yes
<i>blpZ</i>	2.08	1.13	Immunity protein BlpZ	Yes
<i>pncP</i>	1.87	1.13	Putative protease	Yes
<i>JN057_RS03420</i>	1.51	2.13	hypothetical protein	Yes
<i>JN057_RS04230</i>	2.69	1.27	hypothetical protein	Yes
<i>JN057_RS04385</i>	1.76	1.08	IS30 family transposase	No
<i>JN057_RS05275</i>	-1.48	-2.02	helix-turn-helix transcriptional regulator	Yes
<i>piaC</i>	1.06	1.85	iron compound ABC uptake transporter permease	Yes
<i>JN057_RS07945</i>	2.06	1.19	Hydroxyacylglutathione hydrolase	Yes

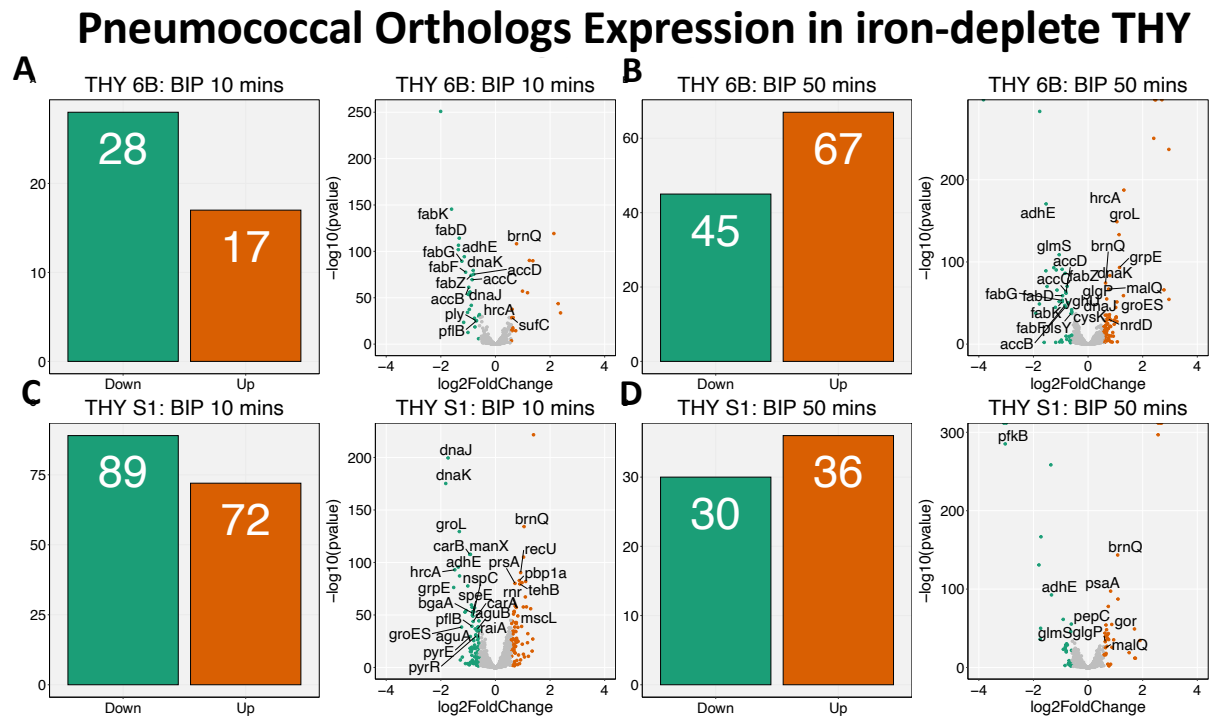
(Table continued)



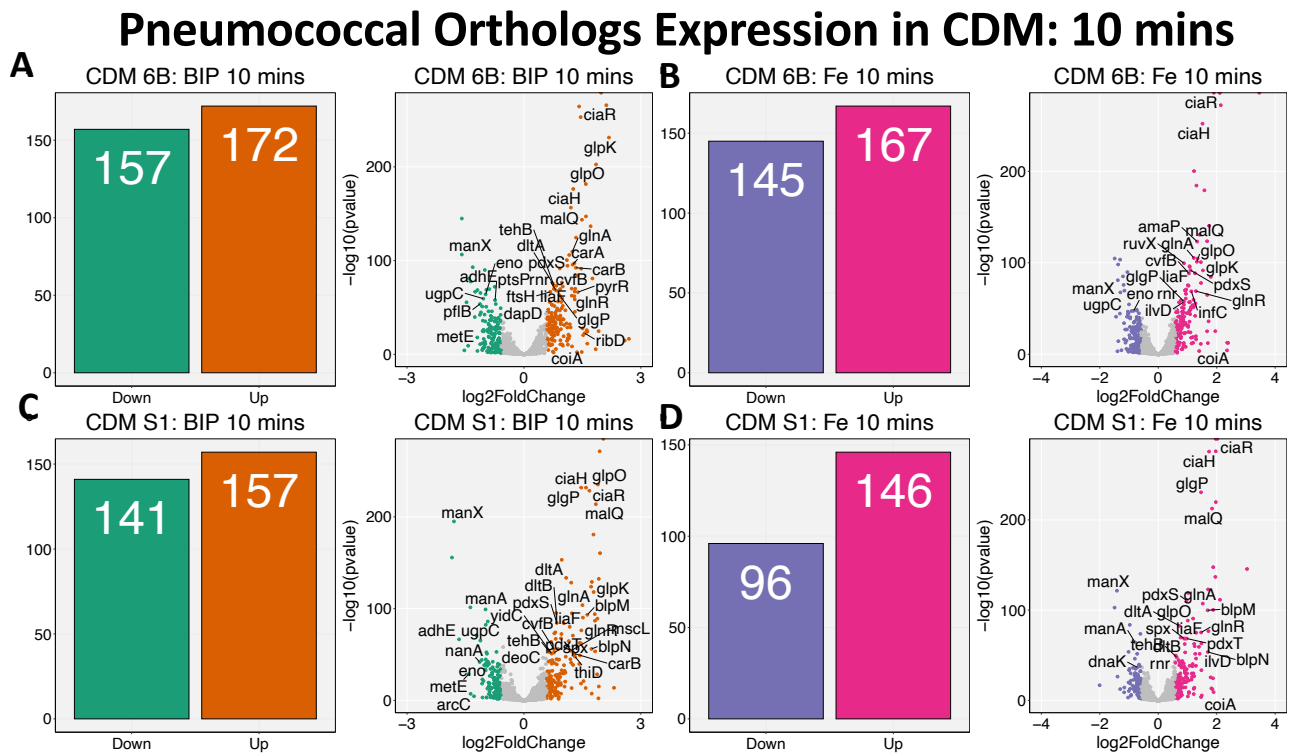
<i>satA</i>	-1.90	-1.33	ABC transporter, putative N-acetylneuraminate-binding protein	Yes
<i>JN057_RS08165</i>	3.76	1.04	YhcH/YjgK/YiaL family protein	Yes
<i>JN057_RS08660</i>	1.40	1.77	Small integral membrane protein	Yes
<i>JN057_RS08710</i>	-1.36	-2.04	PEP phosphonmutase	No
<i>JN057_RS10095</i>	1.65	3.38	MarR family winged helix-turn-helix transcriptional regulator	No
<i>fucL</i>	-1.15	2.26	L-fucose isomerase	Yes
<i>fucK</i>	1.41	2.45	rhamnulokinase	Yes
<i>pavB</i>	1.70	1.06	Cell wall surface anchor family protein	Yes

### 6.3.19 Analysis of *S. pneumoniae* orthologs confirmed strain-specific transcriptional adaptation to iron-deplete and iron-replete conditions

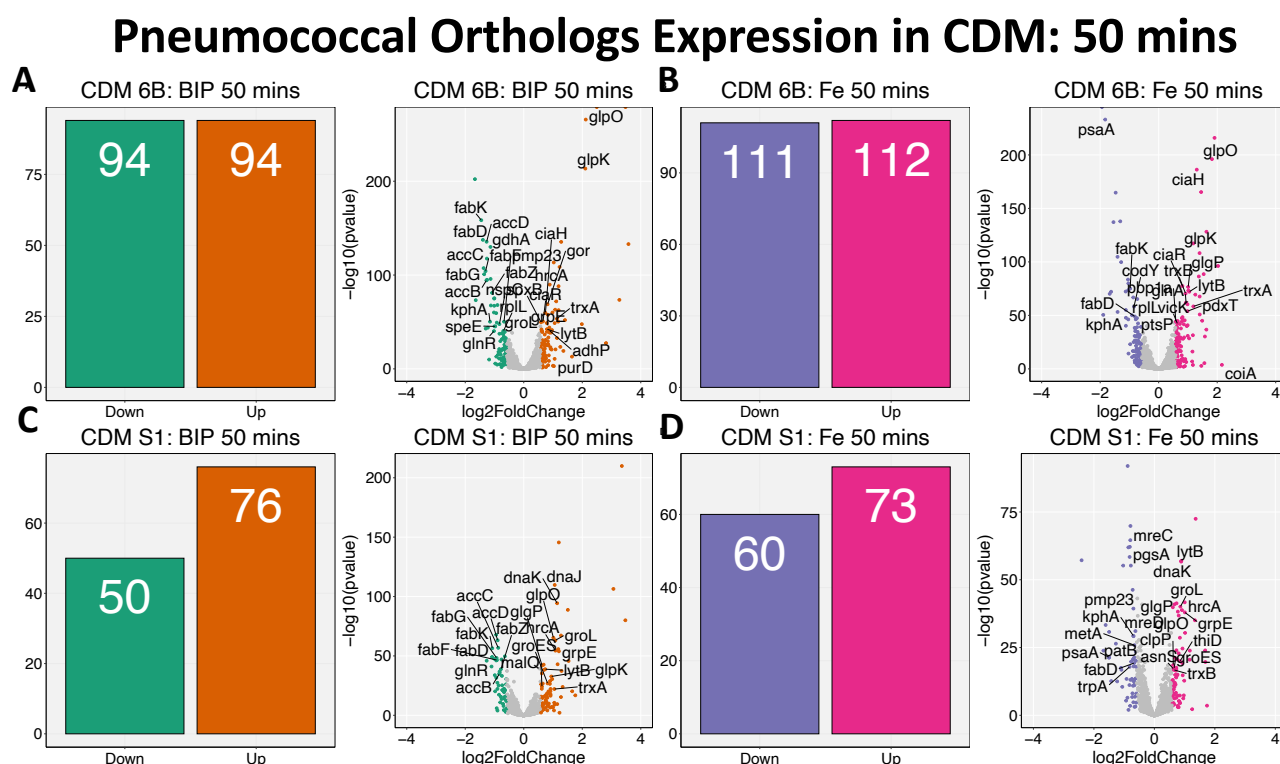
The total number of coding sequences (genes) in *S. pneumoniae* is ~2100 but *S. pneumoniae* are highly plastic with up to 10% of gene content divergent between strains. Comparison of gene contents of our study strains (BHN418UCL and BVJ1JL) relative to reference strains [TIGR4 (serotype 4), D39V (serotype 2), BHN418 (serotype 6B), INV104 (serotype 1), 670-6B (serotype 6B)] indicated 1473 core genes were commonly present in all the strains. Of the total 2338 genes present in the pangenome of BHN418UCL (serotype 6B) and BVJ1JL (serotype 1) strains, 1674 were core between both strains. Differential gene expression analysis of the core genes of serotype 1 (BVJ1JL) and serotype 6B (BHN418UCL) showed similar gene expression patterns to those observed with the complete gene sets of each strain for all experimental conditions (**Figure 6.19**, **Figure 6.20**, and **Figure 6.21**). These results suggest that differences in regulation of core genes of *S. pneumoniae* may be the driver to strain specific adaptation in response to iron availability.



**Figure 6.19: Differential gene expression of 1674 orthologs of serotype 6B (BHN418UCL) and serotype 1 (BVJ1JL) exposed to iron-deplete conditions for 10 minutes (A and B) and 50 minutes (C and D) in THY.** Volcano and bar plots showing up (brown) and down (green) regulated genes by **(A and B)** serotype 6B and **(C and D)** serotype 1 strains. Each point in the volcano plot indicates gene expression with colour indicative of expression direction (brown = upregulated, green = downregulated). Where available in the original file, gene names were annotated on volcano plots, allowing maximum of 40 gene annotations on each plot. DEGs were defined by fold change  $>1.5$  and false discovery rate  $FDR < 0.05$ .






**Figure 6.20: Differential gene expression of 1674 orthologs of serotype 6B (BHN418UCL) and serotype 1 (BVJ11JL) exposed to iron-deplete (A and C) and iron-replete (B and D) conditions for 10 minutes in CDM.** Volcano and bar plots showing up (iron-deplete = brown, iron-replete = pink) and down (iron-deplete = green, iron-replete = pink) regulated genes in **(A and B)** serotype 6B and **(C and D)** serotype 1 strains. Each point in the volcano plot indicates gene expression with colour indicative of expression. Where available in the original file, gene names were annotated on volcano plots, allowing maximum of 40 gene annotations on each plot. DEGs were defined by fold change >1.5 and false discovery rate FDR <0.05.

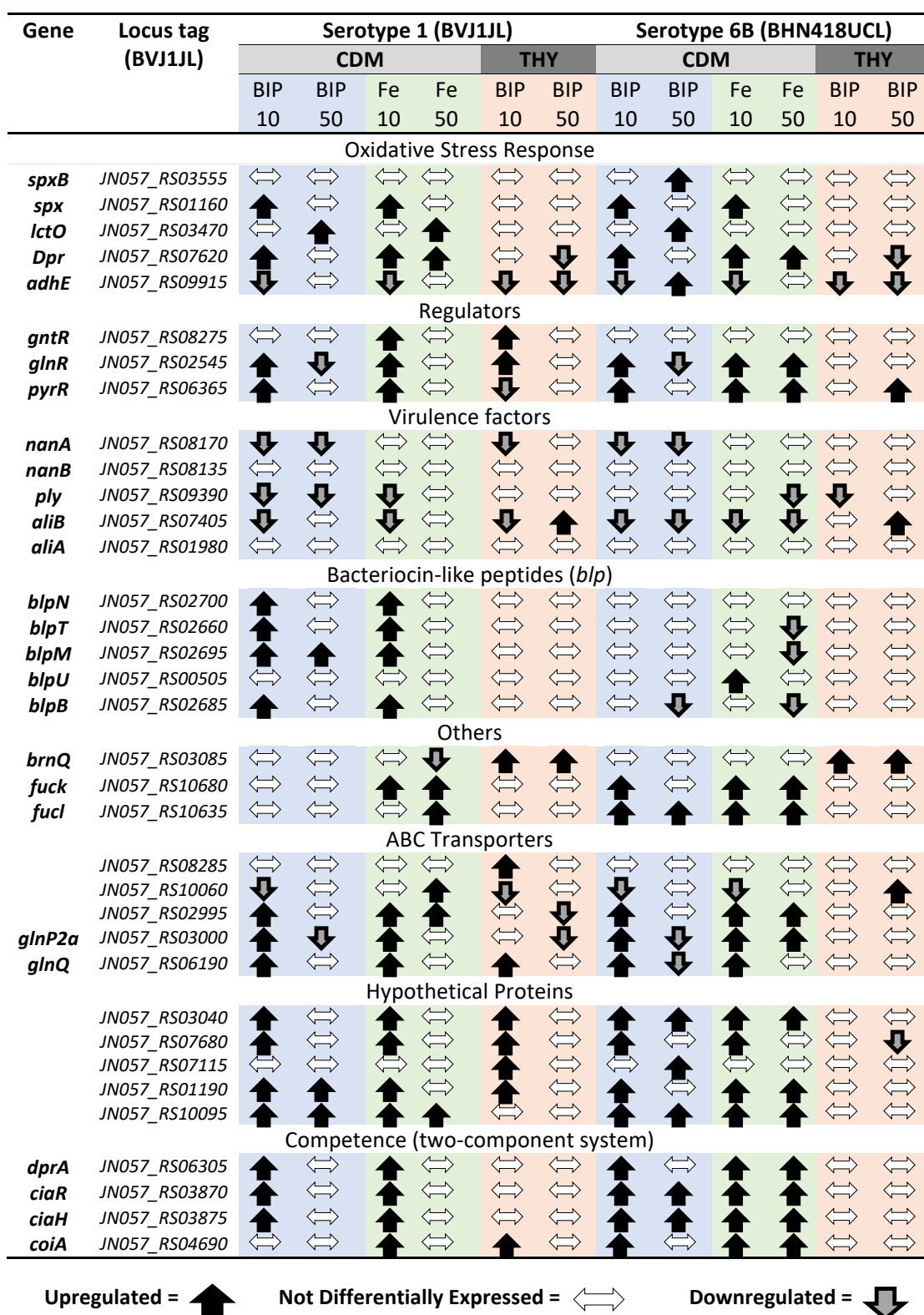


**Figure 6.21: Differential gene expression of 1674 orthologs of serotype 6B (BHN418UCL) and serotype 1 (BVJ1JL) exposed to iron-deplete (A and C) and iron-replete (B and D) conditions for 50 minutes in CDM.** Volcano and bar plots showing up (iron-deplete = brown, iron-replete = pink) and down (iron-deplete = green, iron-replete = pink) regulated genes in **(A and B)** serotype 6B and **(C and D)** serotype 1 strains. Each point in the volcano plot indicates gene expression with colour indicative of expression. Where available in the original file, gene names are annotated on volcano plots, allowing maximum of 40 gene annotations on each plot. DEGs were defined by fold change >1.5 and false discovery rate FDR <0.05.

Gene	Locus tag (BVJ1JL)	Serotype 1 (BVJ1JL)						Serotype 6B (BHN418UCL)					
		CDM				THY		CDM				THY	
		BIP 10	BIP 50	Fe 10	Fe 50	BIP 10	BIP 50	BIP 10	BIP 50	Fe 10	Fe 50	BIP 10	BIP 50
<i>piuA</i>	JN057_RS09020	↔	↔	↓	↓	↑	↑	↔	↔	↓	↓	↑	↑
<i>piuB</i>	JN057_RS09005	↔	↔	↓	↓	↑	↑	↔	↔	↓	↓	↑	↑
<i>piuC</i>	JN057_RS09010	↔	↔	↓	↓	↑	↑	↔	↔	↓	↓	↑	↑
<i>piuD</i>	JN057_RS09015	↔	↔	↓	↓	↑	↑	↔	↔	↓	↓	↑	↑
<i>piaA</i>	JN057_RS05280	↔	↔	↔	↔	↔	↔	↔	↔	↓	↔	↔	↔
<i>piaB</i>	JN057_RS05285	↔	↔	↔	↔	↔	↔	↔	↔	↔	↔	↔	↔
<i>piaC</i>	JN057_RS05290	↔	↔	↔	↔	↔	↔	↑	↔	↑	↔	↔	↔
<i>piaD</i>	JN057_RS05295	↔	↔	↔	↔	↔	↔	↔	↔	↔	↔	↔	↔
<i>ritR</i>	JN057_RS02035	↔	↔	↔	↔	↔	↔	↔	↔	↔	↔	↔	↔
<i>psaA</i>	JN057_RS07965	↔	↔	↔	↓	↑	↑	↔	↔	↔	↓	↔	↔
<i>psaB</i>	JN057_RS07955	↔	↔	↔	↓	↔	↑	↔	↓	↔	↓	↔	↔
<i>psaC</i>	JN057_RS07960	↔	↔	↔	↓	↔	↑	↔	↔	↔	↓	↔	↔
<i>copY</i>	JN057_RS03540	↔	↑	↔	↔	↑	↑	↔	↑	↔	↔	↑	↑
<i>cupA</i>	JN057_RS03545	↔	↑	↔	↔	↑	↑	↔	↑	↔	↔	↑	↑
<i>copA</i>	JN057_RS03550	↔	↑	↔	↔	↑	↑	↔	↑	↔	↔	↑	↑
<i>adcA</i>	JN057_RS10690	↔	↔	↔	↔	↔	↔	↔	↔	↔	↔	↔	↔
<i>adcAll</i>	JN057_RS04805	↑	↔	↑	↔	↔	↔	↔	↔	↔	↔	↔	↔
<i>adcB</i>	JN057_RS10695	↔	↔	↔	↔	↔	↔	↔	↔	↔	↔	↔	↔
<i>adcC</i>	JN057_RS10700	↔	↔	↔	↔	↔	↔	↔	↔	↔	↔	↔	↔
<i>adcR</i>	JN057_RS10705	↔	↔	↔	↔	↔	↔	↔	↔	↔	↔	↔	↔
<i>czcD</i>	JN057_RS08950	↔	↔	↑	↔	↔	↔	↔	↔	↑	↔	↔	↔
<i>sodA</i>	JN057_RS03705	↔	↔	↔	↔	↔	↔	↔	↔	↔	↔	↔	↔
<i>mntE</i>	JN057_RS07530	↔	↔	↔	↔	↔	↔	↔	↔	↔	↔	↔	↔
<i>Gls24</i>	JN057_RS08665	↔	↔	↔	↔	↑	↔	↑	↔	↑	↔	↑	↔
<i>Dpr</i>	JN057_RS07620	↑	↔	↑	↑	↔	↓	↑	↔	↑	↑	↔	↓

Upregulated = 
 Not Differentially Expressed = 
 Downregulated = 

**Figure 6.22:** Transcriptomics profile of previously identified transition metals response genes (Table 6.1) across iron-deplete and iron-replete conditions for *S. pneumoniae* serotype 1 (BVJ1JL) and serotype 6B (BHN418UCL) in the present study. Differentially expressed genes (DEGs) were defined as genes with fold-change >1.5 and false discovery rate (FDR) <0.05.



**Figure 6.23: Transcriptomics profile of a cross-section of *S. pneumoniae* gene classes across iron-deplete and iron-replete conditions for serotype 1 (BVJ1JL) and serotype 6B (BHN418UCL).** Differentially expressed genes (DEGs) were defined as genes with fold-change >1.5 and false discovery rate (FDR) <0.05.

#### 6.4 Discussion

The biological mechanisms that allow certain *S. pneumoniae* strains to be more efficient colonisers of the human nasopharynx or mediate their propensity to cause invasive diseases has not been fully elucidated. *S. pneumoniae* must adapt to the different environmental niches it occupies during its progression from a coloniser of the human nasopharynx to causing invasive pneumococcal disease (IPD) (Weiser, Ferreira, and Paton 2018). Differences in adaptation of strains to environmental exposures within the host may therefore determine differences in strain pathogenesis. Transition metals concentration is one of the main differentiating exposures between the lungs (pneumonia-disease causing niche) which has very low iron concentrations and the upper respiratory track (colonisation niche) which has higher iron levels compared to the lungs, but much lower iron levels compared to the bloodstream (septicaemia-disease causing niche) (McDevitt *et al.* 2011; Honsa, Johnson, and Rosch 2013; van Beek *et al.* 2020). Here, we have described for the first time, the global transcriptomic adaptation of clinical *S. pneumoniae* strains in response to iron-deplete and iron-replete conditions using genome wide RNA sequencing (RNA-Seq). To understand strain specific adaptations to iron availability which may explain differences in *S. pneumoniae* invasive and carriage potential, we compared the transcriptomic response of a serotype 6B strain (BHN418), that colonises the nasopharynx frequently but not a common cause of invasive disease, and a serotype 1 strain (BVJ1JL) that is rarely isolated in the nasopharynx but frequently identified in disease and a common cause of pneumococcal epidemics (Brueggemann *et al.* 2004; Weinberger *et al.* 2010; Kwambana-Adams *et al.* 2016; Leimkugel *et al.* 2005).

The robustness of our *in vitro* experimental model was demonstrated by the validation of several previous findings relating to *S. pneumoniae* response to iron availability. Similar to earlier studies, we showed that low iron concentration impaired growth of *S. pneumoniae*, but growth defects can be rescued by repletion with iron or manganese (Akhter *et al.* 2020; Hoyer *et al.* 2018; van Opijnen and Camilli 2012; Tai, Lee, and Winter 1993), indicating manganese may be able to substitute for biological functions of iron in *S. pneumoniae* (**Figure 6.3A** and **Figure 6.3B**). We also confirm that the expression of pneumococcal iron uptake genes *piuABCD* is directly influenced by iron availability with gene expression significantly upregulated in iron-deplete conditions and repressed in

iron-replete conditions (**Figure 6.22**). The *piuABCD* (originally named *pit1ABCD*) loci together with *piaABCD* (originally named *pit2ABCD*) were the first characterised pneumococcal iron uptake genes which were shown to contribute to *S. pneumoniae* virulence (Brown, Gilliland, and Holden 2001). Leveraging our well controlled experimental design, our data suggests that *piuABCD* loci may be the preferred or dominant iron response loci for clinical *S. pneumoniae* strains. *piuABCD* and *piaABCD* are the main iron import systems in pneumococcus and the deletion of one is able to compensate for the other with severe attenuation of virulence in murine models of pneumoniae and infection observed when both loci were deleted (Brown, Gilliland, and Holden 2001; Brown *et al.* 2001). In their discovery studies of these iron import systems, Brown and colleagues (Brown *et al.* 2002) inferred the *pia* loci to be the dominant loci in iron-deplete conditions based on mRNA expression levels of *pia*, *piu*, and *pit* using real-time PCR. Using the more sensitive and specific RNA-Seq method for gene expression analysis, our data showed that expression of the the *piu* loci, but not for the *pia* loci, is regulated in response to iron levels for clinical strains of serotype 1 and serotype 6B (**Figure 6.22**). The preferential upregulation of *piu* and not *pia* in iron restricted growth environments by *S. pneumoniae* has also been reported for TIGR4 (serotype 4) and D39 strains (serotype 2) using proteomics (Hoyer *et al.* 2018) and microarray transcriptomics methods (Jimenez-Munguia *et al.* 2018). The expression of the conserved *cop* operon genes (*copY*, *copA*, and *cupA*) was also correlated with iron availability for both strains of serotype 1 and serotype 6B (**Figure 6.22**). The *cop* operon was discovered through DNA microarray studies to be transcriptionally induced by copper stress and phenotypic analysis confirmed contribution to colonisation and invasive disease (Shafeeq *et al.* 2011). The genes in this locus were upregulated in both early (10 minutes) and late (50 minutes) timepoints in response to iron depletion by both strains (**Figure 6.22**). The expression of *cop* was however not downregulated in iron-replete conditions, suggesting that iron may not be the preferred substrate of *cop*. Given the sensitivity of *cop* to iron stress in the present study, and the unlikelihood to encounter high nonphysiologically levels of copper *in vivo* (**Figure 6.22**), we postulate that *cop* may play overlapping functionality in response to regulating transition metal concentrations in the host. Consistent with our findings, van Opijnen *et al* identified the *cop* operon to be induced by transition metal stress which is



alleviated by availability of iron or manganese (van Opijnen and Camilli 2012) confirming that the expression of this operon is sensitive to iron levels.

Genes encoding the virulence factor pneumococcal surface adhesins (*psaABC*) showed strain specific and time-dependent expression levels in response to iron-concentrations. While the *psa* locus was downregulated at later (but not early) timepoints for both strains in iron-replete conditions, in iron-deplete conditions only the serotype 1 strain upregulated gene expression of *psaABC* with *psaA* being upregulated in both early and later timepoints (**Figure 6.22**). *psaA* is a *S. pneumoniae* virulence factors that is important for colonisation and plays a key role in import and export of manganese and zinc with manganese being the high affinity substrate for *psaA* (Tseng *et al.* 2002; Honsa, Johnson, and Rosch 2013; Berry and Paton 1996). Given the high correlation of *psaABC* expression in response to iron concentrations found here, *psa* ligands may also be involved in homeostatic regulation of cellular iron levels. Furthermore, the differential regulation of *psaA* between serotype 1 and serotype 6B may contribute to the difference in phenotypes of these strains during colonisation (Berry and Paton 1996).

The specific mechanism by which *S. pneumoniae* senses iron is unknown and *ritR* has been suggested as the orphan master regulator of iron in *S. pneumoniae* (Ulijasz *et al.* 2004). The presumed role of *ritR* was based on DNA microarray studies which indicated that *ritR* was associated with transcription of iron uptake and oxidative stress genes. However, although the *piu* and *cop* operon gene expressions were clearly sensitivity to iron concentrations, expression of the *ritR* gene was not differentially regulated in any of our experimental conditions (**Figure 6.22**). *ritR* is thought to regulate iron toxicity in *S. pneumoniae* by binding the *piu* promoters to prevent iron uptake and was linked to the expression of oxidative stress response genes including *dpr* and *adhE* (Ulijasz *et al.* 2004; Glanville *et al.* 2018). The expression of *ritR* has been previously reported to increase at least two folds in iron rich growth conditions in studies using microarrays hybridisation (Ulijasz *et al.* 2004) and real-time reverse transcriptase PCR (Ong *et al.* 2013) methods. Using RNA-Seq method, our data does not support the involvement of *ritR* in response to iron levels in clinical *S. pneumoniae* strains of serotype 1 and serotype 6B (**Figure 6.22**). It is plausible that the D39 strain which was used for the study of Ulijasz and colleagues

(Ulijasz *et al.* 2004) regulates gene expression of *ritR* differently to the clinical *S. pneumoniae* strains in response to iron. Nevertheless, the correlation of *ritR* expression with a small subset of iron response genes (only two genes of the *piu* loci) in the original study suggest *ritR* may be a poor candidate as master iron regulator in *S. pneumoniae* (Ulijasz *et al.* 2004). Furthermore, unlike iron response genes such as *psa*, *pia* and *piu* loci, which are established pneumococcal virulence factors, *ritR* does not contribute to pneumococcal virulence (Ong *et al.* 2013) indicating that it is dispensable for pneumococcal growth and survival. Similarly, Miao and colleagues (Miao *et al.* 2018) have suggested the general stress response gene *gls24* (*SPD\_1590* in D39) as a possible iron response transporter with functionality similar to *piuA*. We corroborated results of their study by showing that *gls24* is sensitive to low-iron conditions resulting in elevated gene expression in both clinical strains of serotype 1 and serotype 6B (**Figure 6.22**). However, in contrast to *piuA*, *gls24* was also upregulated in iron-replete conditions for serotype 6B. Furthermore, the upregulation of *gls24* by both serotype 1 and serotype 6B strains only occurred in the early timepoints in both iron-deplete and iron-replete conditions. This characteristic is consistent with the behaviour of general response proteins during early adaptation to new environments.

Central to the adaptation of *S. pneumoniae* to environmental perturbation are two-component systems (TCS) which are involved in sensing environment and driving the cellular response (Paterson, Blue, and Mitchell 2006). Unsurprisingly, these TCSs were the core upregulated gene sets in both iron-deplete and iron-replete conditions (**Figure 6-23**). Pathway enrichment analyses of differentially regulated genes in both iron-replete and iron-deplete conditions identified two-component systems, ribosome synthesis, amino acid synthesis, and carbohydrate biosynthesis to be significantly enriched, highlighting broad biological pathways impacted by iron levels. The dysregulation of these pathways by reduction in iron availability to pneumococcus is supported by previous studies (Hoyer *et al.* 2018; Nanduri *et al.* 2008; Jimenez-Munguia *et al.* 2018). We have supplemented these findings by showing that these biological pathways that are also enriched by clinical strains in iron rich environments.

We have identified several conserved genes which significantly altered their gene expression in our dataset and may be putative iron response genes of *S. pneumoniae* (**Figure 6.23**). Examples of such genes include, *JN057\_RS10095*, which is defined as a MER like family protein, *JN057\_RS03040*, defined as a metalloprotease, and *JN057\_RS04930*, defined as surface protein of unknown function. An example of a MER-like family protein in *S. pneumoniae* is the zinc efflux gene *czcD* involved in heavy metal export. We have also identified some genes with broader annotation that showed highly correlated gene expression in response to iron availability such as *brnQ*. *brnQ* is a branched chain amino-acid (BCAA) that upregulated gene expression in iron-deplete THY but not in the iron-replete conditions (**Figure 6.23**). The upregulation of BCAA *liv* during iron limitation has been demonstrated in D39 strain (Hoyer *et al.* 2018). Strain specific difference in gene expression of bacteriocin-like peptide (*blp*) genes which were upregulated in both iron-deplete and iron-replete conditions in CDM for serotype 1 but not for serotype 6B was observed (**Figure 6.23**). *blpRH* is a TCS, and similar to previous studies, bacteriocin upregulation was associated with expression of competence genes which purportedly increases transformation rates (Kjos *et al.* 2016). Bacteriocins are expressed by pneumococcus to outcompete other bacteria. The expression of these genes by serotype 1 in stress conditions may exhibit a competitive advantage during colonisation. However, The expression of bacteriocins by our strains was only observed in CDM and not THY (**Figure 6.23**). In addition, the upregulation of *blp* was only observed in early timepoints and could possibly constitute a general stress response to growth in CDM. Another gene that showed time-dependent gene expression in response to iron availability was *czcD* (**Figure 6.23**).

The current study has some limitations. We classified 10 minutes exposure to iron conditions as “early” timepoints. Bacterial pathogens are known to adapt gene expression quickly following exposure to new stimuli and it is likely that some “early” transcriptomic signals were missed by 10 minutes. Therefore, studying an earlier timepoint like 5 mins post exposure may be ideal for future experiments. Nonetheless, as far as we are aware, the 10 minutes timepoint in this study is the earliest used to investigated *S. pneumoniae* transcriptome in response to iron levels. The amount of iron in our CDM formulation was not measured but is likely to be low since iron was not added to the CDM media

formulation. Thus, the consequences of further chelation of the CDM on pneumococcal transcriptome were unknown. However, we have shown that chelation with BIP does not affect *S. pneumoniae* survival in the short-term as strains remained viable in these environments up to four hours following exposure to the chelating agent (**Figure 6.4**).

In this work, we have described the transcriptome of clinical strains of *S. pneumoniae* in response to iron-deplete and iron-replete conditions, providing an overview of adaptation of pneumococcus to these environments. Several potentially novel iron response genes were identified which will require future studies to validate. Given the importance of transition metals as co-factors for many biological processes that are essential for bacterial survival, the observation that inter-strain transcriptomics differences mainly involved the core genome provides promising avenues for targeting these genes for the development of potential interventions that may not be restricted by serotype.

## Chapter 7 : Transcriptomic analysis of the adaptation of *S. pneumoniae* to changes in temperature

## 7.1 Introduction

A change in environmental temperature is one of the most important conditions that *S. pneumoniae* must adapt to as part of its lifestyle (Klinkert *et al.* 2012; Kwon *et al.* 2003). Transmission of *S. pneumoniae* between individuals is mainly thought to be via aerosol droplets, highlighting the ability of the pathogen to survive in the external environment during transmission (Morimura *et al.* 2021; Bogaert, de Groot, and Hermans 2004; Weiser, Ferreira, and Paton 2018; Zafar *et al.* 2017). During colonisation of the human upper respiratory tract (URT), the natural reservoir, *S. pneumoniae* is exposed to temperatures at the epithelial surfaces of approximately 30°C (Lindemann *et al.* 2002). Following translocation from the nasopharynx to the lung, middle ear, the bloodstream or CSF, *S. pneumoniae* has to adapt to temperatures of approximately 37°C which may reach 40°C as part of the host febrile response (Kwon *et al.* 2003; Pandya *et al.* 2005; Aprianto *et al.* 2018). Microbes mount rapid protective responses following significant changes to ambient temperature which involves coordinated expression of multiple genes (Weiser, Ferreira, and Paton 2018; Klinkert and Narberhaus 2009). Encountering elevated temperature within the mammalian host is a signal for virulence genes expression and rapid synthesis of protective molecules such as heat shock proteins (HSP) (Kwon *et al.* 2003; Chao *et al.* 2020; Mandin and Johansson 2020; Zhang *et al.* 2009; Filannino *et al.* 2018). Bacterial HSPs include chaperones and proteases which are a highly conserved set of proteins that protect microorganisms from adverse environmental effects such as elevated temperatures by primarily controlling correct protein folding to maintain cellular homeostasis (Zhang *et al.* 2009; Kwon *et al.* 2003; Maleki *et al.* 2016). Temperatures outside the host may also have important influence on transmission and survival of *S. pneumoniae* (Jusot *et al.* 2017). For example, in the meningitis belt in sub-Saharan African (Greenwood 1999), where in addition to meningococcal epidemics, frequent pneumococcal meningitis epidemics have been reported, exposure to high temperatures and reduced humidity have been identified as pre-disposing factors (Jusot *et al.* 2017).

In order to mount an appropriate cellular response to temperature fluctuations, bacterial pathogens have evolved thermosensing capabilities (Klinkert *et al.* 2012; Loh *et al.* 2018). RNA thermosensors (RNATs) are a common regulatory principle used by pathogenic bacteria to sense and respond to changing environmental temperatures. RNATs are

usually located at the 5' untranslated region (UTR) of mRNAs and function by masking the ribosomal binding sites (RBS) and blocking downstream translation at low temperatures (e.g., <30°C) (Loh *et al.* 2018; Klinkert *et al.* 2012). At elevated temperatures (e.g., 37°C), the RBS becomes accessible and allow for translation initiation (Loh *et al.* 2018; Klinkert *et al.* 2012). RNATs have been identified early on in a diverse set of pathogenic bacteria including *Listeria monocytogenes* (Johansson *et al.* 2002), *Escherichia coli* (Morita *et al.* 1999), *Yersinia pestis* (Hoe and Goguen 1993), and *Salmonella* (Hurme *et al.* 1997). Bacterial pathogens are thought to use thermosensing capabilities to signal presence in new environments such as mammalian cells which initiates remodelling of its gene regulation including expression of bacterial virulence factors (Klinkert and Narberhaus 2009; Loh *et al.* 2018). In *Salmonella* for example, the SPI-2 pathogenicity island which encodes a type III secretion system (T3SS) is thermoregulated (Ono *et al.* 2005; Duong *et al.* 2007). At temperatures below 30°C, the bacterium is unable to express T3SS, but this is rapidly reversed during growth at elevated temperatures of 37°C (Ono *et al.* 2005; Duong *et al.* 2007). In the mucosal pathogen *N. meningitidis*, the expression of virulence factors necessary for capsule biosynthesis, the expression of factor H binding protein, and sialylation of lipopolysaccharides have been found to be thermoregulated, demonstrating a bacterial mechanism for immune evasion during translocation from URT to disease causing niches (Loh *et al.* 2013). Initial efforts by the same group to identify a similar mechanism for the related mucosal pathogen *S. pneumoniae* were unsuccessful (personal communication). However, recent data from Eichner and colleagues (Eichner *et al.* 2021) suggests that polysaccharide capsule biosynthesis and expression of factor H proteins by *S. pneumoniae* and *Haemophilus influenzae* are also regulated by RNATs, underscoring the importance of adaptation to temperature changes for mucosal pathogens pathogenesis. Similar to *S. pneumoniae*, *N. meningitidis* and *H. influenzae* are asymptomatic co-colonisers of the human URT that can translocate to cause life-threatening diseases (Man, de Steenhuijsen Piters, and Bogaert 2017).

With the aim of describing the regulon of *S. pneumoniae* during adaptation to different temperatures (21°C, 29°C, 33°C, and 40°C compared to 37°C), Panya and colleagues (Pandya *et al.* 2005) utilised microarrays to investigate the gene expression of *S.*

*pneumoniae* TIGR4 strain (serotype 4). The authors identified significant reprogramming of *S. pneumoniae* transcriptome with 29% of genes differentially regulated by temperature change (Pandya *et al.* 2005). Using proteomics, Lee *et al* identified 25 proteins that were differentially expressed following transfers of growth culture of a D39 strain (serotype 2) from temperatures of 37°C to 42°C (Myoung-Ro 2006). Among the *S. pneumoniae* proteins that were found to be upregulated were the HSPs *dnaK*, *greE* and central metabolism genes such as hypoxanthine phosphoribosyl transferase (*hgt*), adenylate kinase (*adk*), and phosphoglycerate mutase (*gpmA*) (Myoung-Ro 2006). More recently, Aprianto *et al* employed the higher resolution RNA sequencing (RNA-Seq) method to study the transcriptome of *S. pneumoniae* D39 strain (serotype 2) in a set of “infection relevant conditions” including temperature change (Aprianto *et al.* 2018). Here, the authors identified a cluster of 19 *S. pneumoniae* genes that were particularly sensitive to increasing temperature and suggested that these are important for translocation from the nasopharynx to other disease-causing niches. This “temperature response” gene cluster included HSPs (*hrcA*, *grpE*, *dnaK*, *dnaJ*, and *groESL*), components of the proteolytic Clp complex (*clpL*, *clpP* and *clpC*), ABC transporters, surface proteins, and a pneumococcal two-component system (**Table 7.1**) (Aprianto *et al.* 2018). Although these previous studies provided an important roadmap for understanding pneumococcal gene expression in response to temperature, some gaps were apparent. Firstly, most of the studies used the laboratory adapted strains D39 (serotype 2) which has been shown to have significant metabolic and gene expression differences when compared to clinical strains that cause disease in humans (Bruckner *et al.* 2004; Lanie *et al.* 2007). Secondly, the evolution of the *S. pneumoniae* transcriptome in response to temperature over time has not been elucidated. We speculate that *S. pneumoniae* exhibit a rapid initial rewiring of its transcriptome in response to temperature which is continually fine-tuned over time. Indeed, understanding transcriptomic evolution following temperature change may provide novel insights into pathogenesis of *S. pneumoniae*. Finally, previous studies have not been designed to allow direct comparison of transcriptomic response of multiple *S. pneumoniae* strains to temperature changes. A hundred *S. pneumoniae* serotypes have been described thus far but the propensity of pneumococcal serotypes to cause IPD is varied with only a subset of strains commonly linked to IPD (Ganaie *et al.* 2020; Brueggemann *et al.* 2004; Weinberger *et al.* 2010). The mechanisms that make certain *S.*



*pneumoniae* serotypes to be more likely to cause IPD is not fully understood but may be related to differences in adaptation of different strains to environmental exposures encountered within the host. We speculate that this may include temperature.

In this chapter, we used RNA-Seq to investigate the transcriptomic adaptation of clinical strains of *S. pneumoniae* following a change in temperature from 30°C (temperature of the nasopharynx) to 37°C (core body temperature) over time. *S. pneumoniae* serotype 1 (BVJ1JL) was used as our model ‘invasive’ strain because it is rarely isolated in the nasopharynx of healthy individuals and is a common cause of IPD and epidemics especially in sub-Saharan Africa (Brueggemann *et al.* 2004; Weinberger *et al.* 2010; Ruckinger *et al.* 2009; Kwambana-Adams *et al.* 2016). We compared the transcriptome of our serotype 1 strain to a serotype 6B strain (BHN418), which is a common coloniser of the nasopharynx in healthy individuals but less frequently a cause of IPD (Brueggemann *et al.* 2004; Weinberger *et al.* 2010).

## 7.2 Methods

### 7.2.1 Conservation of temperature response genes in clinical *S. pneumoniae* strains

In Chapter 3, it was shown that *S. pneumoniae* strains grew exponentially at growth temperatures of 37°C and 39°C, but the strains exhibited slower growth rate at temperatures of 30°C with longer lag phase and lower maximum growth densities. We also observed a trend of strain specific differences in growth at 30°C with invasive strains such as TIGR4 (serotype 4) and BVJ1JL (Serotype 1) growing quicker than carriage strains such as BHN418 (serotype 6B). In their transcriptomic analysis of the D39 strain (serotype 2), Aprianto *et al* (Aprianto *et al.* 2018) identified a set of 19 *S. pneumoniae* genes that were upregulated during exposure to elevated temperatures (**Table 7.1**). The cluster of “temperature responsive” genes were suggested to be important for pneumococcal pathogenesis. Using publicly available clinical genome sequences of *S. pneumoniae*, we investigated the conservation of these temperature responsive genes across clinical *S. pneumoniae* strains. The rationale for conducting this analysis at the time was that given

the importance of adaptation to temperature changes in *S. pneumoniae* lifestyle, the conservation of the then recently described temperature responsive genes by Aprianto *et al* (Aprianto *et al.* 2018) may be associated with differential growth phenotypes observed under experimental conditions in Chapter 3 and possibly the invasiveness of *S. pneumoniae* strains.

### 7.2.2 *S. pneumoniae* genomes used for gene conservation analysis

All genome sequences available were downloaded from the [BigsDB database](#) for *S. pneumoniae* on 29<sup>th</sup> January 2020 (Jolley, Bray, and Maiden 2018). As of this date the total number of genomes was 9,147. All invasive genome sequences for which the source was indicated to be from blood or cerebrospinal fluid (CSF) were downloaded resulting in a total of 511 genomes. The majority of genomes deposited on BigsDB are draft genome sequences that were generated by short-gun sequencing methods and prone to having gaps. We therefore also sought complete *S. pneumoniae* genome assemblies that were closed or assembled *de-novo* from a single contig using long read sequencing for more detailed analysis. The seven complete genome sequences used are summarised in **Table 7.2**.

**Table 7.1:** List of “temperature responsive” genes identified by RNA-Seq using *S. pneumoniae* D39 strain (serotype 2) (Aprianto *et al.* 2018).

Gene/ Protein	Annotation
<i>hrcA</i>	Heat-inducible transcription repressor HrcA
<i>grpE</i>	Heat shock protein GrpE
<i>dnaK</i>	Chaperone protein DnaK
<i>dnaJ</i>	Chaperone protein DnaJ
<i>ctsR</i>	Transcriptional regulator
<i>SPV_2171</i>	Hypothetical protein
<i>groESL</i>	Heat shock protein 60 family co-chaperone

(Table continued)	
<i>clp complex (clpL, clpC, clpP)</i>	Putative ATP-dependent Clp proteinase ATP-binding subunit ClpL
<i>glnQ2</i>	Glutamine ABC transporter ATP-binding protein GlnQ2, degenerate
<i>cbpA</i>	choline binding protein A
<i>SPV 2019-20</i>	pneumococcal two component system
<i>thiXYZ-SPV 2027</i>	thiamin ABC transporter

**Table 7.2:** *S. pneumoniae* complete genome sequences used for gene conservation analysis.

Genome assembly	Sequencing method	Serotype	Sample type	Date sampled	Country
CP027540.1	Hybrid	2 (D39)	Cell culture	1916	USA
NC_003028.3	PacBio RS II	TIGR4	Cell culture	2001	USA
LS483449.1	PacBio RS II	7C	Blood	2007	South Africa
LS483451.1	PacBio RS II	16F	Blood	2010	South Africa
LS483450.1	PacBio RS II	12F	Blood	2011	South Africa
LS483448.1	PacBio RS II	9V	Blood	2006	South Africa
LS483523.1	PacBio RS II	6B	Blood	2005	South Africa

### 7.2.3 Investigation of gene conservation

Local BLAST and custom scripts were used to investigate the presence and conservation of heat response genes using the D39 gene sequences of the strain used by Aprianto *et al* as reference (Aprianto *et al.* 2018). A cut-off of ten mismatches in nucleotide sequence identity was arbitrary selected to determine sequence homology. Strains that have more than 10 nucleotide sequence mismatches were investigated at protein level using MEGA5 (v10.2) to determine whether mutations were synonymous or nonsynonymous. Where mutations were shown to be nonsynonymous, conserved domain analysis was performed using the NCBI conserved domain database to determine whether functional protein domains were affected.

#### 7.2.4 Bacterial strains for transcriptomics experiment

Clinical strains of *S. pneumoniae* serotype 1 (BVJ1JL) and serotype 6B (BHN418) were used to investigate the transcriptomic basis of the growth difference in response to temperature increase. The serotype 1 strain (BVJ1JL) was isolated from the nasopharynx of a 9-year-old child in Malawi in 2015 and we have published its complete genome sequence (Betts *et al.* 2021). The BHN418 strain was isolated from nasopharynx of a child in Sweden between 1997 – 2004 (Browall *et al.* 2014). The BHN418 strain is being used to inoculate healthy adults as part of the controlled Experimental Human Pneumococcal Carriage (EHPC) study in Liverpool (Mitsi *et al.* 2020) and has been re-sequenced and characterised in Chapter 4 as BHN418UCL. Purified frozen stocks of BVJ1JL and BHN418 prepared from the same batch as in Chapter 6 were used here.

#### 7.2.5 Growth media

All bacterial growth in this Chapter was performed in the Nose Mimicking Condition (NMC) chemically defined media (CDM) described by Aprianto *et al.* (Aprianto *et al.* 2018). NMC best supports growth of *S. pneumoniae* strains among the selection of chemically defined media (CDM) tested in Chapter 3 and was used to investigate *S. pneumoniae* transcriptome in iron-deplete and iron-replete conditions in Chapter 6. This CDM has also been used to generate RNA-Seq data for *S. pneumoniae* laboratory strain D39 (serotype 2) in response to temperature change (Aprianto *et al.* 2018). Using the same CDM allows us to compare the transcriptome of D39 strain (serotype 2) more readily to clinical strains of serotype 1 and serotype 6B in response to temperature change. NMC is a modification of the Sicard's defined growth medium (Sicard 1964; Aprianto *et al.* 2018) and discussed in Chapter 3.

#### 7.2.6 Growth experiments

To assess growth kinetics of *S. pneumoniae*, 300µL of bacterial stock was added to 25mL of freshly prepared CDM and 200µL added to sterile flat-bottomed 96-well microtiter plates. Plates were incubated at 30°C with 5% CO<sub>2</sub> in a microplate reader (Spark, Tecan), reading the optical density at 600nm every 30 minutes after 5 seconds orbital shaking

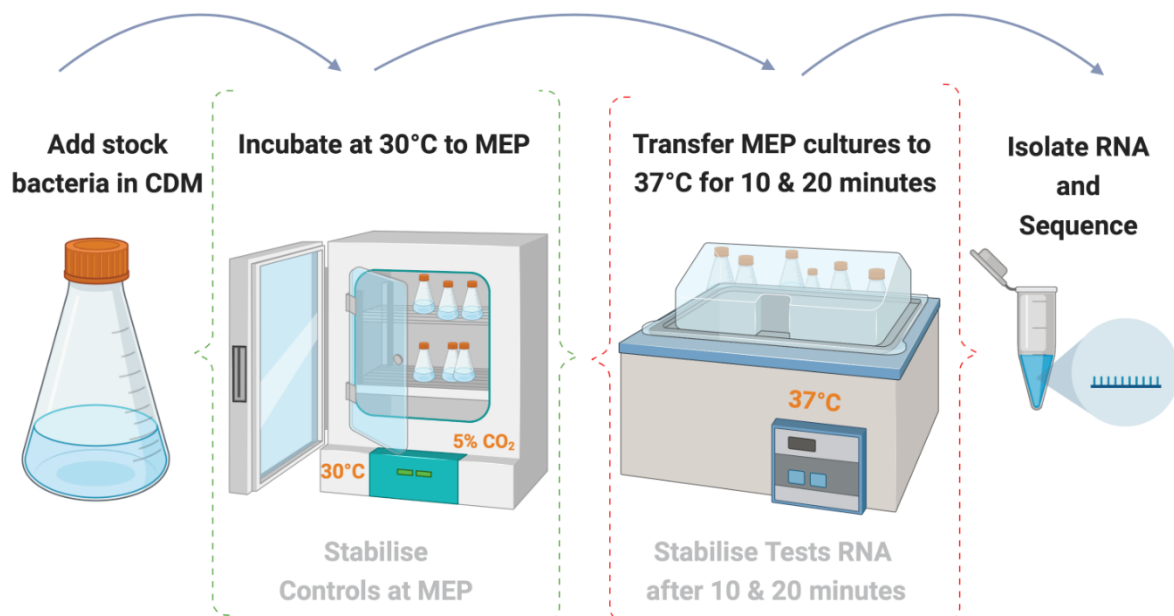
over a 20-hour period. At least four technical replicates were included for each condition and average ODs from three independent experiments were used to analyse bacterial growth. OD readings from control wells containing media but not bacteria, were used to control for background OD (blanking) by subtracting from test well with bacteria. Growth of *S. pneumoniae* was also manually investigated in 250mL glass flasks using the same stock bacterial concentrations used for microplate reader with final culture volume of 25mL. Manual growth was conducted to confirm mid-exponential phase (MEP) of growth in serotype 1 and serotype 6B during growth at 30°C. Manual growth of *S. pneumoniae* in glass flasks was done at 30°C and 5% CO<sub>2</sub> and the number of colonies forming units (CFUs) was assessed at the selected MEPs. ODs for manual growth experiments were measured every hour.

### 7.2.7 Temperature changes transcriptomic experiment design

The transcriptomics protocol used for temperature experiment was similar to iron transcriptomic experiment in Chapter 6, minimising the effects of bacteria growth on the transcriptome by using short exposure times at elevated temperatures. The same CDM formulation used to select MEP was used to set up transcriptomics experiment for *S. pneumoniae* serotype 1 (BVJ1JL) and serotype 6B (BHN418) to minimise growth variance due to different media batches. *S. pneumoniae* strains were grown at 30°C with 5% CO<sub>2</sub> to MEP and transferred to a water bath pre-set at 37°C. Cultures were incubated in a water bath with agitation at 150rpm for either 10 or 20 minutes. At the completion of the incubation times, the Qiagen RNeasy Protect<sup>®</sup> Bacteria Reagent was added to stabilise RNA according to manufacturer's protocol and stored at -80°C. The RNeasy Protect<sup>®</sup> Bacteria Reagent was immediately added to the experimental control samples following growth to MEP point at 30°C. Stabilised RNA was then store at -80°C until RNA isolation. For each experiment condition four biological replicates were included. **Figure 7.1** illustrates the experimental design for the temperature shift from 30°C to 37°C and **Table 7.3** described experimental conditions that were investigated.

**Table 7.3:** Growth conditions to induce *S. pneumoniae* transcriptome for serotypes 1 (BVJ1JL) and serotype 6B (BHN418)

Condition name	Growth description
<b>MEP</b>	<i>S. pneumoniae</i> growth in chemically defined media (CDM) to mid-exponential phase (MEP) at 30°C
<b>Temp10</b>	<i>S. pneumoniae</i> grown in CDM to MEP at 30°C; then transfer to 37°C for 10 minutes
<b>Temp20</b>	<i>S. pneumoniae</i> grown in CDM to MEP at 30°C; then transfer to 37°C for 20 minutes

**Figure 7-1: Illustration of the temperature change transcriptomics experiment.**

*S. pneumoniae* was grown to mid exponential phase at 30°C with 5% CO<sub>2</sub> and then transferred to a water bath pre-set at 37°C and incubated for 10 minutes and 20 minutes. Experiments were conducted with four biological replicates for each condition in different 250mL glass flask. After each timepoint, bacterial transcription was stopped by the addition of RNAprotect (Qiagen) stabilization reagent to all replicates. Stabilized RNA was then stored at -80°C until RNA extraction. Figure was created online using BioRender (Perkel 2020).

### **7.2.8 RNA isolation and quantification**

An optimised RNA isolation protocol based on the commercially available MirVana kit (Thermo Fisher) and described in detail in Chapter 2, was used to isolate RNA from stored experiment samples. RNA isolation, DNase treatment of RNA, and RNA quantification were performed as described in Chapter 6. Ten microliters of the three biological replicates with the best RNA quantification results for each condition were submitted to Vertis Biotechnologie AG (Freising, Germany) for RNA sequencing.

### **7.2.9 Library preparation and sequencing**

#### **7.2.9.1 rRNA depletion and whole transcriptome cDNA preparation**

At Vertis Biotechnologie AG (Freising, Germany), the total RNA samples were examined again by capillary electrophoresis. From the total RNA samples, ribosomal RNA molecules were depleted using an in-house protocol. The ribodepleted RNA samples were first fragmented using ultrasound (2 pulses of 30 seconds each at 4°C). An oligonucleotide adapter was then ligated to the 3' end of the RNA molecules. First-strand cDNA synthesis was performed using M-MLV reverse transcriptase and the 3' adapter as primer. The first-strand cDNA was purified, and the 5' Illumina TruSeq sequencing adapter was ligated to the 3' end of the antisense cDNA. The resulting cDNA was PCR-amplified to about 10-20 ng/μL using a high-fidelity DNA polymerase. The TruSeq barcode sequences which are part of the 5' and 3' TruSeq sequencing adapters were used. The cDNA was purified using the Agencourt AMPure XP kit (Beckman Coulter Genomics) and was analysed by capillary electrophoresis using Shimadzu MultiNA microchip electrophoresis system

#### **7.2.9.2 cDNA from +/- terminator exonuclease treatment of RNA (differential RNA-Seq)**

Differential RNA-Seq (dRNA-Seq) was performed from two pools of total RNAs with equal amounts of the RNA preparations each. One Pool containing serotype 6B samples and the other pool containing serotype 1 samples. First, the total RNA samples were fragmented using ultrasound (4 pulses of 30 seconds at 4°C). This was followed by a treatment with T4 Polynucleotide Kinase (New England Biolabs). The samples were then split into two

halves and one half was subjected to terminator exonuclease treatment (+TEX), the other half was left untreated (-TEX). This was done to enable differentiation of primary transcripts which carry 5'PPP from processed transcripts (such as rRNAs and tRNAs) which carry 5' monophosphate (5'P) and thus facilitate identification of transcript start sites (TSS) and transcript termination sites (TTS) (Sharma and Vogel 2014). The treated RNA samples were poly(A)-tailed using poly(A) polymerase and then, the 5'PPP structures were removed using RNA 5' Polyphosphatase (Epicentre). Afterwards, an RNA adapter was ligated to the 5'-monophosphate of the RNA. First-strand cDNA synthesis was performed using an oligo(dT)-adapter primer and the M-MLV reverse transcriptase. The resulting cDNAs were PCR-amplified to about 10-20 ng/ $\mu$ L using a high-fidelity DNA polymerase. The cDNAs were purified using the Agencourt AMPure XP kit (Beckman Coulter Genomics) and were analysed by capillary electrophoresis.

### **7.2.9.3 Pool generation and size fractionation**

For Illumina NextSeq sequencing, the samples were pooled according to cDNA ratios. The cDNA pool was size fractionated in the size range of 200 – 600bp using a preparative agarose gel. An aliquot of the size fractionated pool was analysed by capillary electrophoresis. The primers used for PCR amplification were designed for TruSeq sequencing according to the instructions of Illumina.

### **7.2.9.4 Illumina sequencing**

The cDNA pool was sequenced on an Illumina NextSeq 500 system using 75 bp read length.

### **7.2.10 Sequencing reads quality control, and mapping, transcript counting**

Reads quality control, transcript mapping, and generation of transcripts counts table were replicated as described in Chapter 6.

### **7.2.11 Differential gene expression analysis**

Normalisation and differential gene expression analysis were performed using DESeq2 Bioconductor package (v1.32.0) as described in Chapter 6. The group considerations for



transcriptomic comparisons for the present study are illustrated in Figure (**Figure 7.2**). Unless otherwise stated, transcripts with fold change (FC) >1.5 and false discovery rate (FDR) < 0.01 relative to the control samples were considered as significant differentially expressed genes (DEGs).

### **7.2.12 Statistical analysis**

All statistical analysis and visualizations were performed using latest Bioconductor packages in R environment (v4.1.3). T-test analysis of counts data was performed on DESeq2 normalised counts. Adjusted P-values <0.05 between group comparisons were considered significant.

### **7.2.13 Orthologs and pathway analyses**

Orthologs and pathways analysis were conducted as described in Chapter 6 (Section 6.2.13 – 6.2.14). The orthologs of KEGG annotated strains of 670-6B (serotype 6B) and INV104 (serotype 1) were used for BHN418UCL (serotype 6B) and BVJ1JL (serotype 1) respectively to identify enriched pathways. The Bioconductor R package clusterProfiler (Yu, Wang *et al.* 2012) was used for pathways analysis using BHN418UCL and BVJ1JL orthologs present in the KEGG annotated strains (Kanehisa 2000). Biological pathways with FDR <0.05 were considered enriched for selection of genes.

## Transcriptomic Comparisons: Temperature Shift

2 x 2

Strains	Time Points
Serotype 1	10 mins
Serotype 6B	20 mins

**Figure 7.2: Illustration showing transcriptomic comparisons for temperature experiment.** Transcriptomic data was generated for *S. pneumoniae* serotype 1 (BVJ1JL) and serotype 6B (BHN418) strains following a change from growth at 30°C to elevated temperatures of 37°C. For both strains, *S. pneumoniae* transcriptomic adaptation at 10 minutes and 20 minutes was evaluated following initial growth to mid exponential phase (MEP) at 30°C in a chemically defined medium (CDM). Each experiment was conducted with four biological replicates.

### 7.3 Results

#### 7.3.1 Differences in growth of *S. pneumoniae* strains at different temperatures was not explained by genetic differences in temperature-responsive genes

The conservation of heat response genes in 511 clinical genome sequences of *S. pneumoniae* was high with 14/19 of the genes present in greater than 68% of genome assemblies (**Figure 7.3**). The high polymorphism of *cbpA* gene observed was consistent with previous findings (Iannelli, Oggioni, and Pozzi 2002). However, some known highly conserved pneumococcal genes such as the heat shock protein *groES* showed high dissimilarity between clinical strains and D39 reference sequence at nucleotide level. We suspect that the low conservation of *groES* may be due to poor genome assemblies or

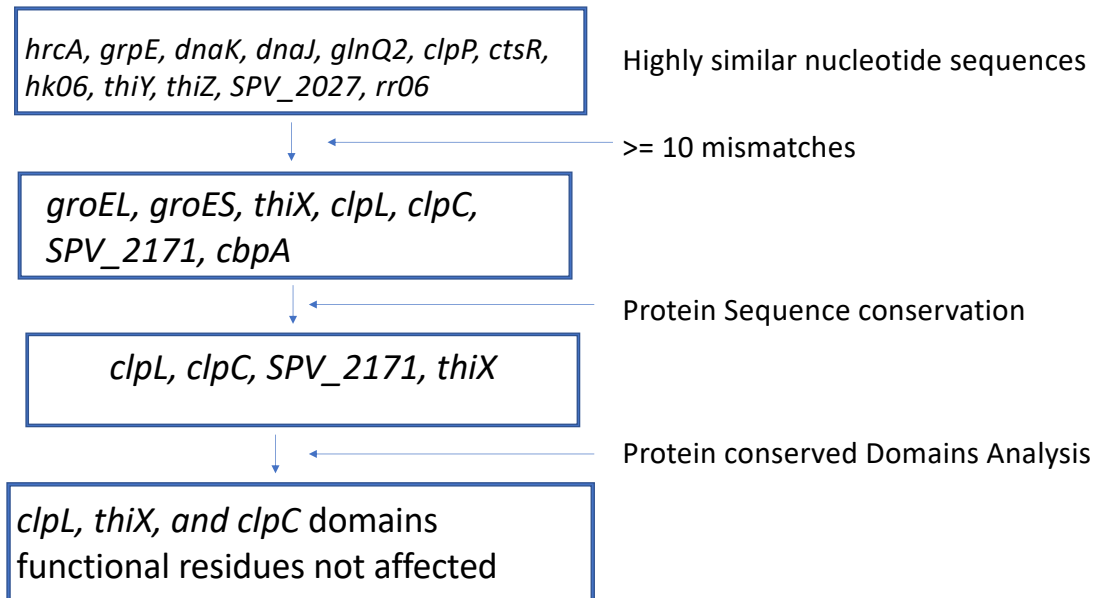
sequencing artifacts leading to frameshift mutations, rather than substantial genetic differences between strains. To investigate this further, the genome sequence of seven high quality *S. pneumoniae* genomes that were sequenced by long read methods and assembled *de-novo* as complete genome sequences were used (**Table 7.2**).

Homologs of twelve (63%) of the temperature responsive genes (*hrcA*, *grpE*, *dnaK*, *dnaJ*, *glnQ2*, *clpP*, *ctsR*, *SPV\_2019*, *thiY*, *thiZ*, *SPV\_2027*, *SPV\_2020*) were present in all seven complete genome sequences. Six of the genes (*groEL*, *groES*, *thiX*, *clpL*, *clpC*, *SPV\_2171*) had greater than 10 mismatches but analysis of protein sequence indicated nonsynonymous polymorphisms for only four of these (*clpL*, *clpC*, *thiX*, *SPV\_2171*). Further analysis of the conserved protein domains of these four genes indicated that the amino acid changes did not affect critical protein residues and thus functional properties of the genes likely remained intact (**Figure 7.4**). Genetic difference of the hypothetical gene *SPV\_2171* was observed between serotype 6B and TIGR4 (serotype 4) genome sequences (**Figure 7.5**). While the *SPV\_2171* gene in serotype 6B was identical to the reference D39 sequences, it was a pseudogene in TIGR4 indicated by the presence of stop codons in the coding sequence in TIGR4 (**Figure 7.5**). *SPV\_2171* was also a pseudogene in the genomes of clinical strains of serotype 12F and serotype 16F and was absent in the genome of serotype 7C. The result of this analysis suggested that *S. pneumoniae* phenotypic growth differences at different temperatures are not explained solely by differences in these previously described temperature response genes. Initial plans to compare the conservation of the “temperature response” genes in 500 invasive and 500 non-invasive clinical *S. pneumoniae* genomes was therefore abandoned.

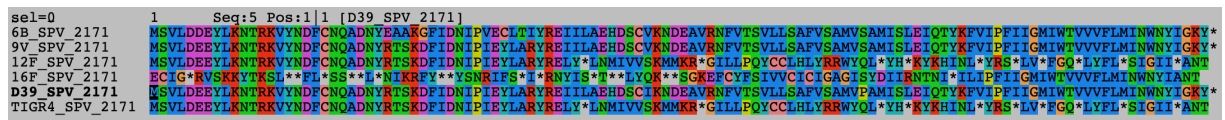
Gene	Presence	% of Positive Samples	Allelic Variants
<i>clpL</i>	26	5.1	12
<i>groES</i>	109	21.3	2
<i>SPV_2027</i>	120	23.5	4
<i>cbpA</i>	219	42.9	152
<i>thiX</i>	219	42.9	13
<i>SPV_2171</i>	352	68.9	23
<i>dnaJ</i>	355	69.5	20
<i>thiZ</i>	438	85.7	20
<i>hrcA</i>	456	89.2	22
<i>cstR</i>	484	94.7	17
<i>clpP</i>	504	98.6	4
<i>groEL</i>	505	98.8	20
<i>glnQ2</i>	508	99.4	43
<i>grpE</i>	508	99.4	9
<i>hk06</i>	510	99.8	46
<i>thiY</i>	510	99.8	8
<i>dnaK</i>	511	100	16
<i>rr06</i>	511	100	16
<i>clpC</i>	511	100	10

**Figure 7.3: Conservation 19 heat response gene in 511 clinical *S. pneumoniae* genome sequences.** Green colours genes have high conservation in genomes investigated while pink colour indicate genes conserved in less than 50% of genomes.

## Protein sequence and function analysis



**Figure 7.4:** Flowchart results of analysis of 19 heat response genes in genomes of 7 clinical strains of *S. pneumoniae* assembled *de-novo* by long read sequencing.



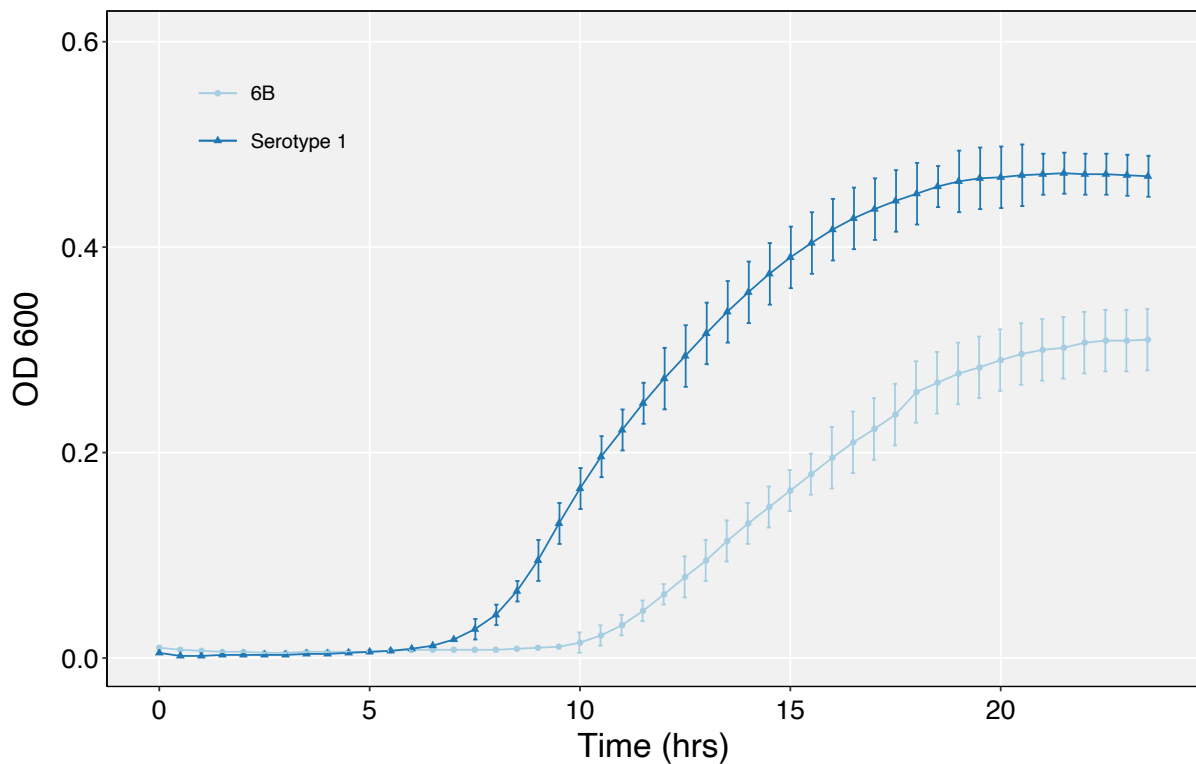
**Figure 7.5:** Protein sequence alignment of hypothetical D39 gene *SPV\_2171* in clinical *S. pneumoniae* strains. The gene was absent in the genome of serotype 7C and was a pseudogene in strains of TIGR4, serotype 16F, and serotype 12F.

### 7.3.2 Growth characteristics of *S. pneumoniae* at 30°C

The growth kinetics of serotype 1 (BVJ1JL) and serotype 6B (BHN418) were evaluated in CDM to identify MEP of the strains in growth at 30°C. The results corroborated earlier findings in Chapter 3 with serotype 1 growing faster than serotype 6B at 30°C (**Figure 7.6**). Given the difference in growth of the strains, MEP growth points selected for transcriptomics experiments were different. The selected MEP points were  $OD_{600} = 0.15$  ( $\pm 0.02$ ) for serotype 6B and  $OD_{600} = 0.23$  ( $\pm 0.02$ ) for serotype 1. The time to reach these growth points was calculate as 13.5 hrs ( $\pm 45$  mins) for serotype 6B and 12 hrs ( $\pm 39$  mins) for serotype 1 when incubated in 250mL glass flasks at 30°C and 5% CO<sub>2</sub>. The number of

colony forming units (CFU) at these harvest points were  $2.9 \times 10^7$  cfu/mL for serotype 6B and  $3.95 \times 10^7$  cfu/mL for serotype 1. Since the starting bacterial count for both strains was identical, the observation that the CFU/mL values at chosen MEP points were within the same log was taken to indicate that both strains were harvested at similar growth phase at the selected ODs.

## Growth of serotype 1 and 6B at 30°C



**Figure 7.6: Growth of *S. pneumoniae* serotypes 1 (BVJ1JL) and serotype 6B (BHN418) in CDM at 30°C.** Data was generated from plate reader incubated at 30°C with 5% CO<sub>2</sub> combining data from independent experiments (n = 3). The error bars represent SEM between the independent experiments.

### 7.3.3 RNA quantification

Using the MEP defined in section 7.3.2, the temperature shift transcriptomic experiment was conducted as illustrated in **Figure 7.1** and detailed in **Table 7.3**. Quantification results of RNA was of high quality with average RNA integrity score of 8.9 (7.7 – 9.5). The minimum RNA yield from any of the four biological replicates in all conditions was 84ng/μL while the maximum was 386ng/μL, values which are adequate for RNA sequencing. The RNA quantification results are shown in **Table 7.4**.

### 7.3.4 Total sequencing reads and mapping

A total of 472,546,979 reads were generated from Illumina sequencing run with an average of ~ 14 million reads (8.9M – 17.5M) for each library. Following removal of poor-quality reads, adapter sequences, and ribosomal reads, the average number of mRNA reads that mapped to the reference genomes was approximate 8M (7.1M – 10.0M) for each library (**Table 7.5**). The mapping results from all libraries is summarised in **Table 7.5**. The average number of ribosomal RNA reads from the libraries was consistent with previous studies that used the Vertis inhouse ribodepletion strategy (Kroger *et al.* 2013; Aprianto *et al.* 2018; Canals *et al.* 2019) (**Table 7.5**). Pool libraries for dRNA-Seq do not require rRNA depletion and this explains the low level of mRNA for these libraries (prefix with “Pool”) in **Table 7.5** (Sharma and Vogel 2014).

**Table 7.4:** RNA quantification results of all replicates of serotypes 1 (BVJ1JL) and serotype 6B (BHN418) under each studied condition for the temperature change experiment. RNA quantification data was generated with Nanodrop and Bioanalyzer.

Serotype	Condition	Rep	Nanodrop (Thermo)			Bioanalyzer (Agilent)	
			ng/ $\mu$ L	260/280	260/230	ng/ $\mu$ L	RIN
6B	MEP	1	135.6	2.13	0.96	91	8.7
6B	MEP	2	151.7	2.11	2.22	97	9.3
6B	MEP	3	126.8	2.11	1.99	89	NA
6B	MEP	4	131.8	2.11	2.25	68	8.9
6B	Temp10	1	132.9	2.12	2.16	90	8.9
6B	Temp10	2	101.2	2.1	2.42	72	9.3
6B	Temp10	3	112.2	2.1	0.9	86	9.1
6B	Temp10	4	105.2	2.08	0.86	64	9.3
6B	Temp20	1	128.4	2.12	1.79	92	9.4
6B	Temp20	2	166.9	2.12	2	106	9.5
6B	Temp20	3	162.8	2.12	2.3	114	9.4
6B	Temp20	4	161.1	2.12	1.3	131	9.5
1	MEP	1	217.9	2.12	2.28	170	8.8
1	MEP	2	287.5	2.14	2.4	231	8.2
1	MEP	3	245.9	2.13	2.25	196	9.4
1	MEP	4	84.1	2.08	2.2	69	NA
1	Temp10	1	110.9	2.08	2.25	91	9
1	Temp10	2	288.6	2.15	1.52	283	9.5
1	Temp10	3	200.8	2.12	2.33	173	9.4
1	Temp10	4	210.4	2.13	2.15	191	9.4
1	Temp20	1	108.8	2.01	1.61	91	NA
1	Temp20	2	206.3	2.14	2.2	185	9.3
1	Temp20	3	301.4	2.15	2.04	287	9.5
1	Temp20	4	386.4	2.14	2.25	284	7.7



**Table 7.5:** Mapping results of RNA-Seq data to *S. pneumoniae* reference genome sequences BHN418UCL (serotype 6B) and BVJ1JL (serotype 1) showing the amounts of rRNA sequences, number of mapped, and unmapped mRNA reads of each library.

Sample	Reference Strain	All (%)	rRNA (%)	Unmapped (%)	Multimap (%)	Assigned (%)
6B_MEP_2	BHN418UCL	11905463 (100)	2128552 (17.88)	556179 (4.67)	101275 (0.85)	8161355 (68.55)
6B_MEP_3	BHN418UCL	13874698 (100)	3716350 (26.79)	646449 (4.66)	101631 (0.73)	8330531 (60.04)
6B_MEP_4	BHN418UCL	13931705 (100)	3193303 (22.92)	659511 (4.73)	110751 (0.80)	8864169 (63.63)
6B_MEP10_1	BHN418UCL	11768817 (100)	1631786 (13.87)	675640 (5.74)	108696 (0.92)	8543880 (72.60)
6B_MEP10_2	BHN418UCL	17093401 (100)	5689279 (33.28)	987729 (5.78)	135261 (0.79)	9201465 (53.83)
6B_MEP10_3	BHN418UCL	14744098 (100)	3610079 (24.49)	807846 (5.48)	135789 (0.92)	9079808 (61.58)
6B_MEP20_2	BHN418UCL	11147333 (100)	1291702 (11.59)	589842 (5.29)	104368 (0.94)	8383019 (75.20)
6B_MEP20_3	BHN418UCL	11141955 (100)	2287301 (20.53)	576672 (5.18)	100963 (0.91)	7405611 (66.47)
6B_MEP20_4	BHN418UCL	10988475 (100)	2266115 (20.62)	579916 (5.28)	103523 (0.94)	7197141 (65.50)
S1_MEP_1	BVJ1JL	11878280 (100)	2147688 (18.09)	231703 (1.95)	59588 (0.50)	8409369 (70.80)
S1_MEP_2	BVJ1JL	11531593 (100)	1907251 (16.54)	213227 (1.85)	56996 (0.49)	8383954 (72.70)

(Table Continued)

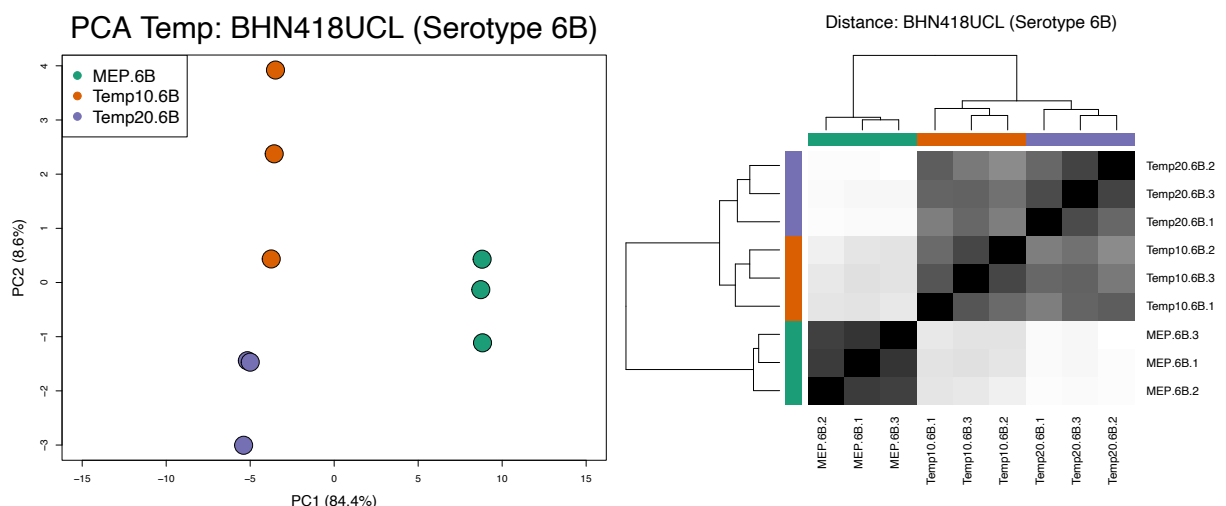
## Chapter 7: Transcriptomic response to temperature change

S1_MEP_3	BVJ1JL	9860627 (100)	1351507 (13.71)	174356 (1.77)	49757 (0.51)	7426040 (75.31)
S1_MEP10_2	BVJ1JL	12505633 (100)	1775036 (14.19)	262827 (2.10)	77265 (0.62)	9459695 (75.64)
S1_MEP10_3	BVJ1JL	10594897 (100)	1256562 (11.86)	249707 (2.36)	79022 (0.75)	8177706 (77.19)
S1_MEP10_4	BVJ1JL	9209006 (100)	752655 (8.17)	176608 (1.92)	61106 (0.66)	7521165 (81.67)
S1_MEP20_2	BVJ1JL	8735944 (100)	757911 (8.68)	191040 (2.18)	59244 (0.68)	7065501 (80.88)
S1_MEP20_3	BVJ1JL	12901009 (100)	1387978 (10.76)	423995 (3.29)	79967 (0.62)	10069819 (78.01)
S1_MEP20_4	BVJ1JL	12489210 (100)	1546527 (12.38)	284548 (2.28)	73508 (0.59)	9625815 (77.07)
Pool-S1	BVJ1JL	27321118 (100)	23058246 (84.40)	230038 (0.84)	17700 (0.07)	3827608 (14.01)
Pool-S1-TEX	BVJ1JL	28255873 (100)	21570955 (76.34)	417288 (1.48)	41562 (0.15)	5337455 (18.89)
Pool-6B	BHN418UCL	29680273 (100)	24663587 (83.09)	419174 (1.41)	31133 (0.10)	4354563 (14.67)
Pool-6B-TEX	BHN418UCL	28303849 (100)	21458235 (75.81)	570868 (2.02)	51454 (0.18)	5304444 (18.74)

### 7.3.5 RNA-Seq data quality

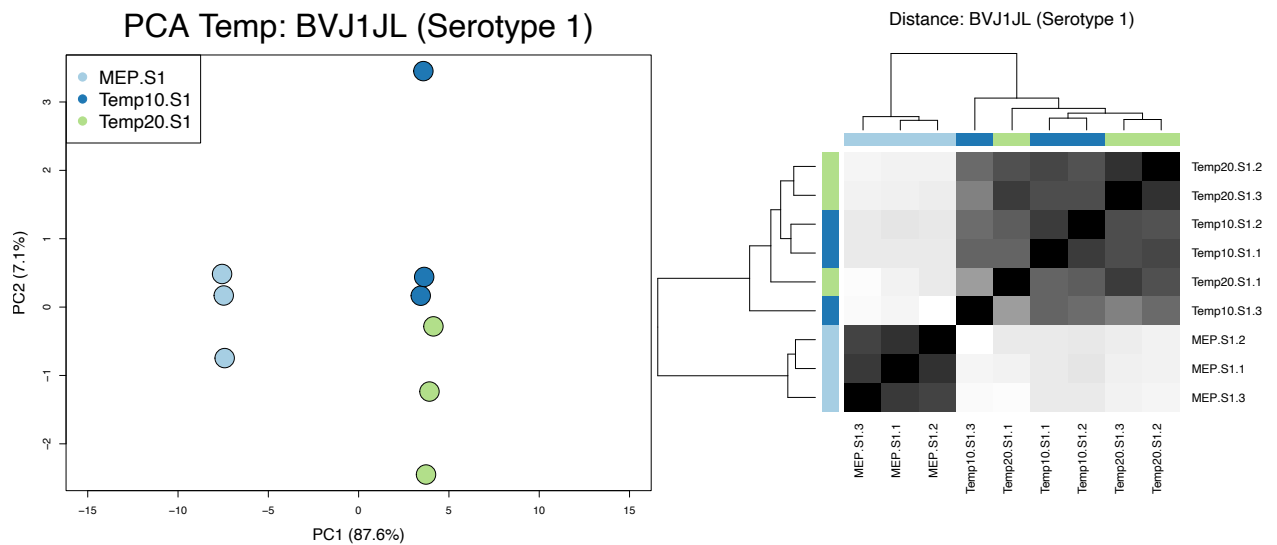
Clustering analysis of the log transformed count data in DESeq2 indicated high quality transcriptomic data with highly correlated biological replicates (**Figure 7.7A** and **Figure 7.7B**). For both serotype 1 and serotype 6B, control samples grown to MEP at 30°C clustered distantly from their respective temperature shifted test samples. Although distance clustering clearly separated biological replicates that were shifted for 10 minutes and 20 minutes for serotype 6B (**Figure 7.7A**), clustering for serotype 1 was less pronounced between the 10 and 20 minutes timepoints (**Figure 7.7B**), suggesting a more closely related transcriptomes for the serotype 1 strain at 10 and 20 minutes.

## Clustering of BHN418UCL: Temperature Shift



**Figure 7.7A: Principal component analysis (PCA) and distance clustering of RNA-seq data showed high reproducibility of biological replicates for *S. pneumoniae* serotype 6B (BHN418) in temperature change experiment.** Replicates of each experimental condition are coloured uniquely (green = controls; brown = temperature change for 10 minutes; purple = temperature change for 20 minutes).

## Clustering of BVJ1JL : Temperature Shift



**Figure 7.7B: Principal component analysis (PCA) and distance clustering of RNA-seq data showed high reproducibility of biological replicates for *S. pneumoniae* Serotype 1 (BVJ1JL) in temperature change experiment.** Replicates of each experimental condition are coloured uniquely (light blue = controls; dark blue = temperature change for 10 minutes; light green = temperature change for 20 minutes).

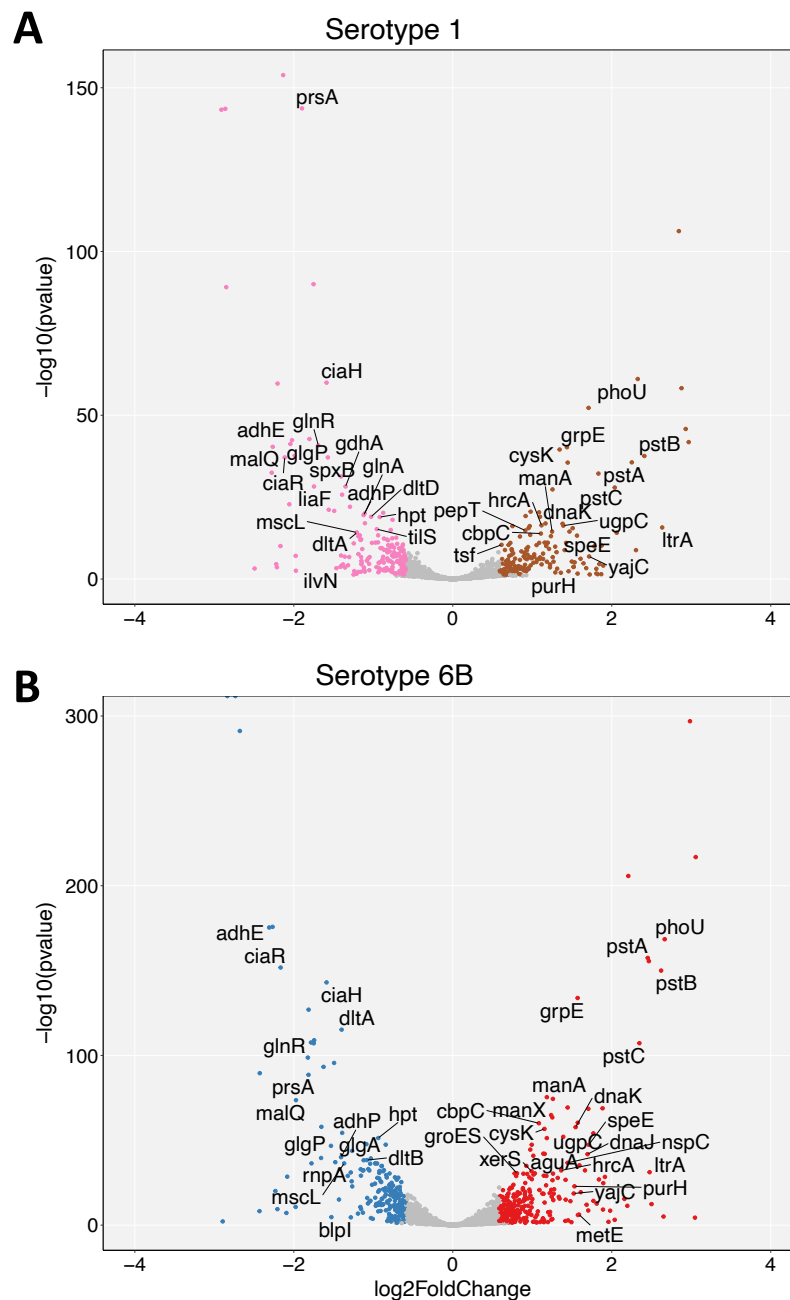
### 7.3.6 Transcriptomics response of *S. pneumoniae* during temperature change from 30°C to 37°C for 10 minutes

During the first 10 minutes following temperature shift from 30°C to 37°C, *S. pneumoniae* differentially regulated a wide range of genes with a total of 451 and 325 genes differentially regulated by serotype 6B and serotype 1 respectively (**Figure 7.8A** and **Figure 7.8B**). While more genes were upregulated than downregulated by serotype 6B, the opposite was true for serotype 1 (**Figure 7.8A**). About 50% (255/517) of the total number of genes differentially regulated at 10 minutes were commonly regulated by both serotype 1 and serotype 6B. In both strains, genes involved in phosphate metabolism were among the most significantly upregulated (**Table 7.6A** and **Table 7.6B**). Unsurprisingly, both strains upregulated several well-known heat shock proteins including *dnaK*, *dnaJ*, *clpC*, *hrcA*, and *grpE*. Genes involved in diverse metabolic processes such as purine metabolism and several hypothetical proteins of unknown function were also upregulated underscoring the broad effect of temperature change on *S. pneumoniae* adaptation (**Table 7.6A** and **Table 7.6B**). However, the expressions of certain genes were strain specific with serotype 6B uniquely upregulating 192 genes while serotype 1 upregulated 70 unique genes **Table 7.6B**. Sixty seven percent (128/192) of the genes that were uniquely regulated by serotype 6B were upregulated. The majority of the uniquely regulated genes by serotype 6B (83%) have orthologs present in serotype 1 but were not differentially regulated by serotype 1. Of the seventy genes that were uniquely upregulated by serotype 1 after 10 minutes at elevated temperature, forty (57%) were upregulated. Forty-one (59%) of the uniquely regulated genes by serotype 1 are present in serotype 6B but were not differentially regulated. Uniquely upregulated genes of serotype 6B that are present in serotype 1 but not differentially regulated included several carbohydrate metabolism and transport genes (e.g., *JN057\_RS02375*, *JN057\_RS02450*, *lacB*, *JN057\_RS01450*, *JN057\_RS01455*, *JN057\_RS01600*, *JN057\_RS01760*), several hypothetical proteins (e.g., *JN057\_RS06005*, *JN057\_RS00635*, *JN057\_RS07000*, *JN057\_RS03740*), some kinases (e.g., *thiM*, *coaA*, *JN057\_RS05345*, *hk06*, *JN057\_RS01745*), several ABC transporters (e.g., *patA*, *JN057\_RS05475*, *JN057\_RS05675*, *JN057\_RS06985*, *JN057\_RS05475*, *JN057\_RS00965*), surface proteins (e.g., *pspC*, *cbpF*), transcription regulators (e.g., *JN057\_RS10530*, *JN057\_RS09170*, ), formate--tetrahydrofolate ligase (*fhs*), and a serine hydrolase (*JN057\_RS07040*). In contrast, the

core genes that were uniquely upregulated by serotype 1 included hypothetical proteins (e.g., *JN057\_RS02415*, *JN057\_RS04280*), peptidase T (*pepT*), trehalose operon repressor (*treR*), a transcriptional repressor *JN057\_RS06980*, a phosphoglycerate mutase (*gpmA*), and glutamine ABC transporter *glnp3a*. Commonly downregulated genes by both strains included competence genes (e.g., *ciaRH*), virulence gene *aliA*, and oxidative stress response genes *adh* and *spxB* (**Table 7.6C** and **Table 7.6D**).

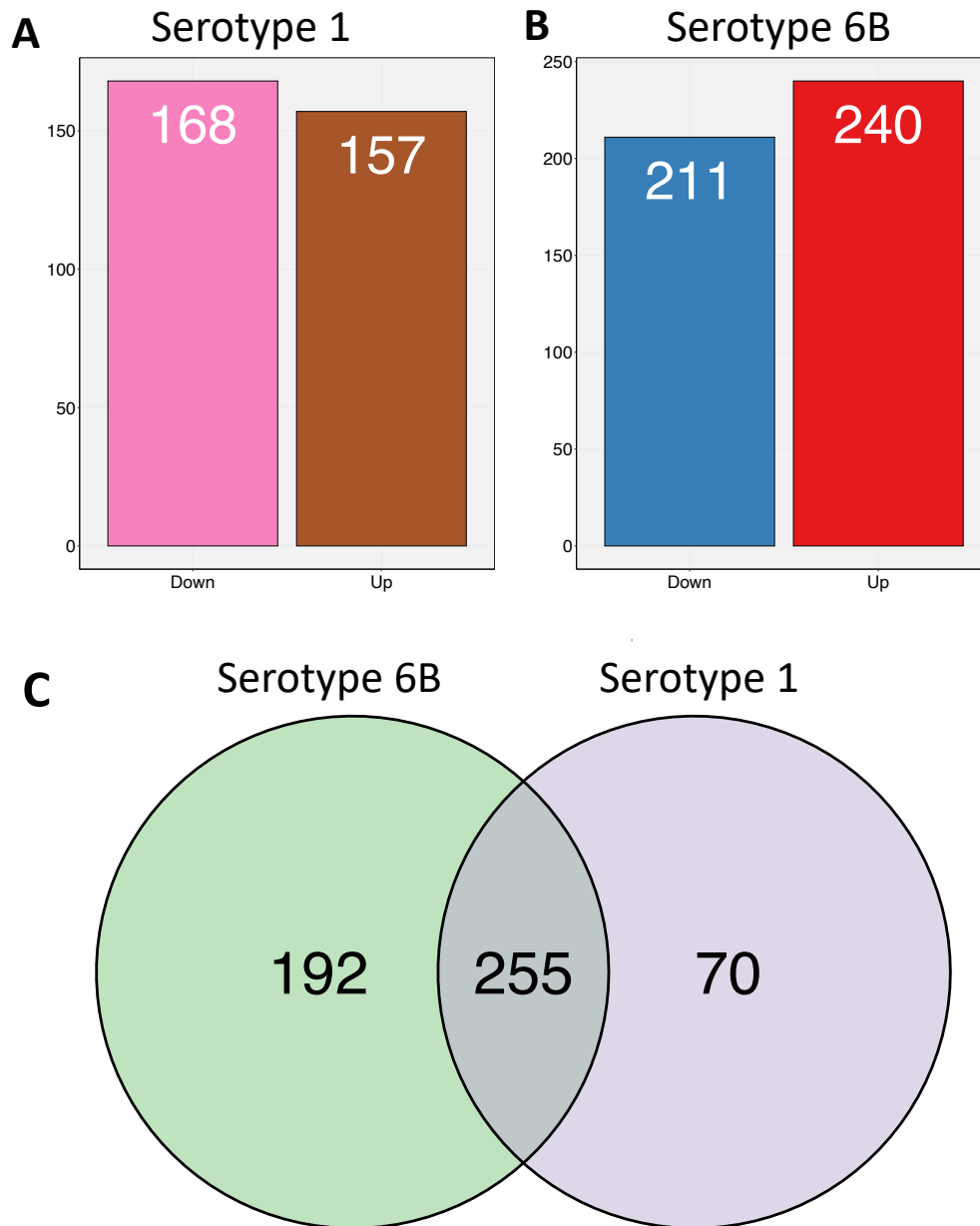
The high temperature requirement A (*htrA*) gene was the third most significantly downregulated gene by both strains (**Table 7.6C** and **Table 7.6D**). Based on bacterial mutagenesis experiments, *htrA* has been previously suggested as a pneumococcal virulence factor with *htrA* deficient strains attenuated in pneumonia and bacteraemia models of infection (Ibrahim *et al.* 2004). In other bacterial pathogens such as *S. pyogenes*, *htrA* is essential for bacterial resistance to thermal stress (Jones *et al.* 2001). However, the suggested regulator of *htrA* in *S. pneumoniae* is the two-component system genes *ciaRH* (Sebert *et al.* 2002; Mascher *et al.* 2003) which in our experiments were among the most significantly downregulated genes in both serotype 1 and serotype 6B (**Table 7.6C** and **Table 7.6D**). Importantly, the previously generated data by Aprianto *et al.* (Aprianto *et al.* 2018) also showed that *htrA*, *ciaR*, and *ciaH* gene expression were downregulated at elevated temperatures for the D39 strain (serotype 2). Apart from the Aprianto *et al.* paper, the work in this thesis is currently the only other RNA-Seq evaluation of the transcriptomic response of *S. pneumoniae* to temperature change.

# DGE at 10 mins



**Figure 7.8A: Volcano plots showing differential gene expression (DGE) by *S. pneumoniae* serotype 1 (BVJ1JL) and serotype 6B (BHN418) following temperature change from 30°C to 37°C for 10 minutes.** Each point on the plot represents a gene with the up- (brown = serotype 1; red = serotype 6B) and down-regulated (pink = serotype 1; blue = serotype 6B) genes indicated by colour. Where available in the original annotation file, short gene names are annotated on volcano plots. DEGs were defined by fold change >1.5 and false discovery rate FDR <0.05.

# DGE at 10 mins



**Figure 7.8B: Total number of differentially expressed genes by of *S. pneumoniae* serotype 1 (BVJ1JL) and serotype 6B (BHN418) following temperature change from 30°C to 37°C for 10 minutes. (A – B) Bar plots show the total number of up- (brown = serotype 1; red = serotype 6B) and down- regulated (pink = serotype 1; blue = serotype 6B) genes. (C) The Venn diagram shows the total number of genes that were shared or uniquely regulated by serotype 1 and serotype 6B. DEGs were defined by fold change >1.5 and false discovery rate FDR <0.05.**



**Table 7.6A:** Top 40 differentially upregulated genes by *S. pneumoniae* serotype 6B (BHN418) following temperature shift from 30°C to 37°C for 10 minutes. Genes were ranked by statistical significance (P-adj). Top upregulated genes in both strains are in bold. RefSeq locus tags of the published sequence of BVJ1JL are shown (Betts *et al.* 2021).

locus	Gene	FC	P-adj	Annotation
<b>Serotype 6B (BHN418)</b>				
<b>JN057_RS10280</b>	<b>pstS1</b>	7.91	1.16E-297	Phosphate ABC transporter periplasmic phosphate-binding protein PstS1
<b>JN057_RS01840</b>	<b>clpL</b>	8.32	1.31E-217	ATP-dependent Clp protease ATP-binding subunit
<b>JN057_RS07405</b>	<b>aliB</b>	4.62	1.77E-206	peptide ABC transporter substrate-binding protein
<b>JN057_RS10300</b>	<b>phoU</b>	6.35	3.37E-169	phosphate signalling complex protein PhoU
<b>JN057_RS10290</b>	<b>pstA</b>	5.47	3.47E-158	phosphate ABC transporter permease PstA
<b>JN057_RS05700</b>	<b>gapN</b>	5.53	2.79E-156	NADP-dependent glyceraldehyde-3-phosphate dehydrogenase
<b>JN057_RS10295</b>	<b>pstB</b>	6.15	1.04E-150	phosphate ABC transporter ATP-binding protein PstB
<b>JN057_RS02615</b>	<b>grpE</b>	2.97	1.41E-134	nucleotide exchange factor GrpE
<b>JN057_RS10285</b>	<b>pstC</b>	5.09	6.27E-108	phosphate ABC transporter permease subunit PstC
<b>JN057_RS02990</b>		2.27	3.89E-76	ferric reductase-like transmembrane domain-containing protein
<b>JN057_RS03575</b>	<b>manA</b>	2.39	4.10E-75	mannose-6-phosphate isomerase, class I
<b>JN057_RS08330</b>	<b>scrR</b>	2.72	5.00E-70	LacI family DNA-binding transcriptional regulator
<b>JN057_RS04390</b>	<b>cad</b>	3.69	1.46E-69	aminotransferase class I/II-fold pyridoxal phosphate-dependent enzyme
<b>JN057_RS08320</b>	<b>scrA</b>	3.26	2.84E-69	sucrose-specific PTS transporter subunit IIBC
<b>JN057_RS01605</b>	<b>manX</b>	2.36	1.41E-65	PTS sugar transporter subunit IIB

(Table continued)

# Chapter 7: Transcriptomic response to temperature change

<b>JN057_RS10060</b>	<b><i>adh</i></b>	2.38	3.26E-64	zinc-dependent alcohol dehydrogenase family protein
<i>JN057_RS08325</i>	<i>scrB</i>	2.97	4.46E-61	glycoside hydrolase family 32 protein
<b>JN057_RS02040</b>	<b><i>cbpC</i></b>	2.12	9.15E-61	choline-binding protein CbpC
<b>JN057_RS02620</b>	<b><i>dnaK</i></b>	2.92	1.85E-58	molecular chaperone DnaK
<b>JN057_RS10875</b>	<b><i>cysK</i></b>	2.22	1.96E-57	cysteine synthase A
<i>JN057_RS04405</i>	<i>Lys1</i>	3.41	8.02E-55	saccharopine dehydrogenase family protein
<b>JN057_RS07660</b>	<b><i>ugpC</i></b>	2.62	9.29E-53	sn-glycerol-3-phosphate ABC transporter ATP-binding protein UgpC
<b>JN057_RS03055</b>		2.27	5.28E-52	amino acid ABC transporter substrate-binding protein
<i>group_614</i>		1.99	4.17E-48	hypothetical protein
<i>JN057_RS04400</i>	<i>speE</i>	3.27	4.44E-48	polyamine aminopropyltransferase
<b>JN057_RS01725</b>	<b><i>basA</i></b>	1.97	3.21E-45	glutathione peroxidase
<i>JN057_RS10535</i>		2.20	7.77E-43	Membrane protease family protein
<i>JN057_RS06300</i>		2.23	1.04E-42	Integral membrane protein
<i>JN057_RS02630</i>	<i>dnaJ</i>	3.24	1.51E-42	molecular chaperone DnaJ
<b>JN057_RS02170</b>		2.02	1.28E-41	carboxymuconolactone decarboxylase family protein
<i>JN057_RS04410</i>	<i>nspC</i>	2.72	1.87E-37	carboxynorspermidine decarboxylase
<i>JN057_RS07005</i>		3.02	6.02E-36	U32 family peptidase
<b>JN057_RS07495</b>		1.90	1.22E-35	NAD(P)H-dependent oxidoreductase
<i>JN057_RS06005</i>		3.17	5.50E-33	hypothetical protein
<b>JN057_RS02610</b>	<b><i>hrcA</i></b>	2.57	5.53E-33	heat-inducible transcriptional repressor HrcA
<i>JN057_RS05825</i>	<i>xerS</i>	1.99	1.82E-32	tyrosine recombinase XerS
<i>JN057_RS10740</i>	<i>ltrA</i>	5.56	6.43E-32	group II intron reverse transcriptase/maturase
<i>JN057_RS09320</i>	<i>groES</i>	1.75	2.90E-31	co-chaperone GroES
<i>JN057_RS10685</i>	<i>fucR</i>	2.05	3.04E-31	L-fucose operon activator
<i>JN057_RS10810</i>	<i>thiZ</i>	1.72	3.59E-31	ABC transporter ATP-binding protein

**Table 7.6B:** Top 40 differentially upregulated genes by *S. pneumoniae* serotype 1 (BVJ1JL) following temperature shift from 30°C to 37°C for 10 minutes. Genes were ranked by statistical significance (P-adj). Top upregulated genes in both strains are in bold. RefSeq locus tags of the published sequence of BVJ1JL are shown (Betts *et al.* 2021).

locus	Gene	FC	P-adj	Annotation
<b>Serotype 1 (BVJ1JL)</b>				
<b>JN057_RS10280</b>	<b>pstS1</b>	7.18	6.00E-107	Phosphate ABC transporter periplasmic phosphate-binding protein PstS1
<b>JN057_RS10300</b>	<b>phoU</b>	5.02	8.46E-62	phosphate signalling complex protein PhoU
<i>group_1371</i>		7.35	5.43E-59	unannotated protein
<i>group_324</i>		3.27	5.59E-53	amino acid ABC transporter ATP-binding protein
<b>JN057_RS01840</b>	<b>clpL</b>	7.62	1.54E-46	ATP-dependent Clp protease ATP-binding subunit
<b>JN057_RS05700</b>	<b>gapN</b>	7.82	1.57E-42	NADP-dependent glyceraldehyde-3-phosphate dehydrogenase
<b>JN057_RS02615</b>	<b>grpE</b>	2.71	5.71E-41	nucleotide exchange factor GrpE
<b>JN057_RS10875</b>	<b>cysK</b>	2.54	2.86E-40	cysteine synthase A
<b>JN057_RS10295</b>	<b>pstB</b>	5.32	2.74E-38	phosphate ABC transporter ATP-binding protein PstB
<b>JN057_RS10290</b>	<b>pstA</b>	4.76	2.38E-36	phosphate ABC transporter permease PstA
<b>JN057_RS07495</b>		2.72	2.81E-36	NAD(P)H-dependent oxidoreductase
<b>JN057_RS10060</b>	<b>adh</b>	3.56	6.35E-33	zinc-dependent alcohol dehydrogenase family protein
<b>JN057_RS10285</b>	<b>pstC</b>	4.10	1.17E-28	phosphate ABC transporter permease subunit PstC
<i>group_1478</i>		2.39	4.38E-28	unannotated protein
<b>JN057_RS03055</b>		1.97	2.69E-21	amino acid ABC transporter substrate-binding protein
<i>JN057_RS04065</i>		2.12	4.24E-21	BMP family protein
<b>JN057_RS02170</b>		1.89	5.86E-20	carboxymuconolactone decarboxylase family protein

(Table continued)

# Chapter 7: Transcriptomic response to temperature change

<i>JN057_RS01050</i>		2.14	9.94E-20	MptD family putative ECF transporter S component
<i>JN057_RS04280</i>		2.24	1.09E-17	hypothetical protein
<b><i>JN057_RS02620</i></b>	<b><i>dnaK</i></b>	2.61	1.31E-17	molecular chaperone DnaK
<b><i>JN057_RS02610</i></b>	<b><i>hrcA</i></b>	2.17	4.12E-17	heat-inducible transcriptional repressor HrcA
<b><i>JN057_RS07660</i></b>	<b><i>ugpC</i></b>	2.62	5.10E-17	sn-glycerol-3-phosphate ABC transporter ATP-binding protein UgpC
<i>JN057_RS08615</i>		1.96	6.80E-17	hypothetical protein
<i>JN057_RS10215</i>		1.68	7.05E-17	glucose-6-phosphate isomerase
<i>JN057_RS01845</i>	<i>ltrA</i>	6.21	1.83E-16	group II intron reverse transcriptase/maturase
<i>JN057_RS05910</i>		1.94	2.85E-16	phosphocarrier protein HPr
<b><i>JN057_RS02990</i></b>		2.85	3.02E-16	ferric reductase-like transmembrane domain-containing protein
<i>JN057_RS04835</i>	<i>pepT</i>	1.88	1.56E-15	peptidase T
<i>JN057_RS09645</i>		2.77	3.19E-15	hypothetical protein
<b><i>JN057_RS03575</i></b>	<b><i>manA</i></b>	2.38	3.28E-15	mannose-6-phosphate isomerase, class I
<b><i>JN057_RS07405</i></b>	<b><i>aliB</i></b>	4.18	7.50E-15	peptide ABC transporter substrate-binding protein
<b><i>JN057_RS02040</i></b>	<b><i>cbpC</i></b>	2.16	1.32E-14	choline-binding protein CbpC
<i>JN057_RS10540</i>		1.96	3.38E-14	PTS ascorbate transporter subunit IIC
<b><i>JN057_RS04390</i></b>	<b><i>cad</i></b>	2.95	5.61E-14	aminotransferase class I/II-fold pyridoxal phosphate-dependent enzyme
<i>JN057_RS10335</i>		2.35	8.38E-14	rhomboid family intramembrane serine protease
<i>JN057_RS03830</i>		1.80	1.04E-13	membrane protein
<b><i>JN057_RS01605</i></b>	<b><i>manX</i></b>	2.23	7.58E-12	PTS sugar transporter subunit IIB
<i>JN057_RS02625</i>		2.29	8.17E-12	hypothetical protein
<b><i>JN057_RS01725</i></b>	<b><i>basA</i></b>	2.12	8.92E-12	glutathione peroxidase
<i>JN057_RS01075</i>	<i>nrdI</i>	2.26	9.10E-12	class Ib ribonucleoside-diphosphate reductase assembly flavoprotein NrdI

**Table 7.6C:** Top 20 differentially downregulated genes by *S. pneumoniae* serotype 6B (BHN418) following temperature shift from 30°C to 37°C for 10 minutes. Genes were ranked by statistical significance (P-adj). Top genes in both strains are in bold. RefSeq locus tags of the published sequence of BVJ11JL are shown (Betts *et al.* 2021).

locus	Gene	FC	P-adj	Annotation
<b>Serotype 6B (BHN418)</b>				
<b>JN057_RS04930</b>		-7.14	0.00E+00	DUF1002 domain-containing protein
<b>JN057_RS03040</b>		-6.66	0.00E+00	M57 family metalloprotease
<b>JN057_RS11045</b>	<b>htrA</b>	-6.40	6.35E-292	trypsin-like peptidase domain-containing protein
<b>JN057_RS08280</b>		-4.81	1.72E-176	ABC transporter ATP-binding protein
<b>JN057_RS09915</b>	<b>adhE</b>	-4.96	4.07E-176	bifunctional acetaldehyde-CoA/alcohol dehydrogenase
<b>JN057_RS03870</b>	<b>ciaR</b>	-4.49	1.60E-152	two-component system response regulator
<b>JN057_RS03875</b>	<b>ciaH</b>	-3.01	9.24E-144	two-component system sensor histidine kinase
<b>JN057_RS08285</b>		-3.52	1.10E-127	membrane protein
<b>JN057_RS10725</b>	<b>dltA</b>	-2.64	6.49E-116	D-alanine--poly(phosphoribitol) ligase subunit
<b>JN057_RS01080</b>		-3.35	1.09E-109	divalent metal cation transporter
<b>JN057_RS02545</b>	<b>glnR</b>	-3.44	2.37E-108	transcriptional repressor GlnR
<b>JN057_RS01980</b>	<b>aliA</b>	-3.36	7.18E-108	peptide ABC transporter substrate-binding protein
<b>JN057_RS11050</b>		-3.54	1.90E-99	ParB/RepB/Spo0J family partition protein
<b>JN057_RS06185</b>	<b>glnHP5</b>	-2.82	2.64E-96	ABC transporter substrate-binding protein/permease
<b>JN057_RS04865</b>	<b>asd</b>	-3.09	5.52E-94	aspartate-semialdehyde dehydrogenase
<b>JN057_RS08275</b>	<b>gntR</b>	-5.38	2.88E-90	GntR family transcriptional regulator
<b>JN057_RS04705</b>	<b>prsA</b>	-3.52	2.63E-89	peptidylprolyl isomerase PrsA
<b>JN057_RS10395</b>	<b>malQ</b>	-3.92	2.30E-74	4-alpha-glucanotransferase
<b>JN057_RS06075</b>	<b>nirC</b>	-3.14	1.16E-58	formate/nitrite transporter family protein
<b>JN057_RS01185</b>	<b>folC</b>	-2.62	4.26E-55	bifunctional folylpolyglutamate synthase/dihydrofolate synthase

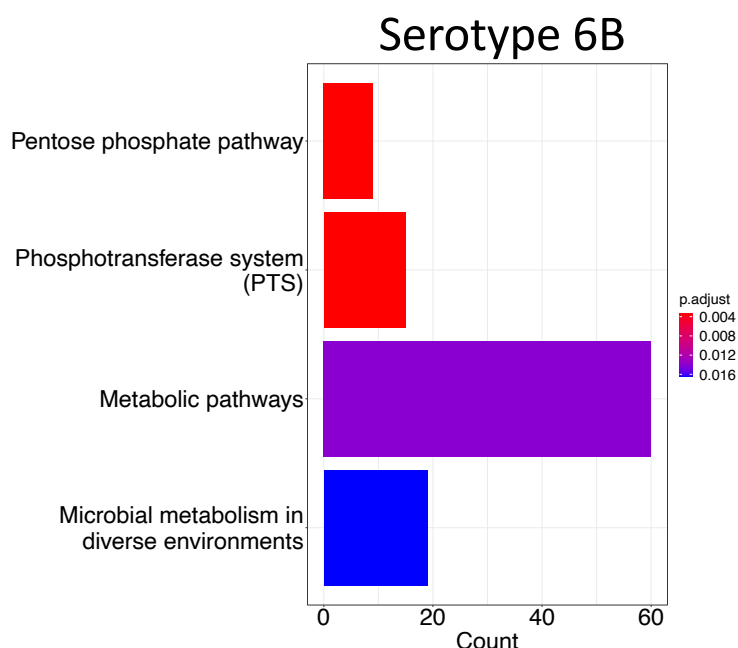
**Table 7.6D:** Top 20 differentially downregulated genes by *S. pneumoniae* serotype 1 (BVJ1JL) following temperature shift from 30°C to 37°C for 10 minutes. Genes were ranked by statistical significance (P-adj). Top genes in both strains are in bold. RefSeq locus tags of the published sequence of BVJ1JL are shown (Betts *et al.* 2021).

locus	Gene	FC	P-adj	Annotation
<b>Serotype 1 (BVJ1JL)</b>				
<b>JN057_RS11050</b>		-4.39	1.42E-154	ParB/RepB/Spo0J family partition protein
<b>JN057_RS04705</b>	<b>prsA</b>	-3.72	2.03E-144	peptidylprolyl isomerase PrsA
<b>JN057_RS11045</b>	<b>htrA</b>	-7.27	2.74E-144	trypsin-like peptidase domain-containing protein
<b>JN057_RS04930</b>		-7.52	5.06E-144	DUF1002 domain-containing protein
<b>JN057_RS04360</b>		-3.37	9.57E-91	PspC domain-containing protein
<b>JN057_RS03040</b>		-7.21	7.63E-90	M57 family metalloprotease
<b>JN057_RS03875</b>	<b>ciaH</b>	-3.01	1.12E-60	two-component system sensor histidine kinase CiaH
<b>JN057_RS01980</b>	<b>aliA</b>	-4.61	2.09E-60	peptide ABC transporter substrate-binding protein
<b>JN057_RS04230</b>		-3.49	1.76E-43	hypothetical protein
<b>JN057_RS09915</b>	<b>adhE</b>	-4.06	4.08E-43	bifunctional acetaldehyde-CoA/alcohol dehydrogenase
<b>JN057_RS01080</b>		-4.12	5.83E-42	divalent metal cation transporter
<b>JN057_RS02545</b>	<b>glnR</b>	-3.23	1.61E-41	transcriptional repressor GlnR
<b>JN057_RS10395</b>	<b>malQ</b>	-4.81	4.27E-41	4-alpha-glucanotransferase
<b>JN057_RS10390</b>	<b>glgP</b>	-3.97	2.24E-38	glycogen/starch/alpha-glucan family phosphorylase
<b>JN057_RS03870</b>	<b>ciaR</b>	-4.33	7.47E-38	two-component system response regulator
<b>JN057_RS08285</b>		-2.97	7.47E-38	membrane protein
<b>JN057_RS08275</b>	<b>gntR</b>	-4.85	3.43E-33	GntR family transcriptional regulator
<b>JN057_RS03555</b>	<b>spxB</b>	-2.65	5.04E-32	pyruvate oxidase
<b>JN057_RS04865</b>	<b>asd</b>	-3.35	5.67E-29	aspartate-semialdehyde dehydrogenase
<b>JN057_RS06490</b>	<b>gdhA</b>	-2.55	5.67E-29	NADP-specific glutamate dehydrogenase

### **7.3.7 Biological pathways enriched by *S. pneumoniae* during temperature shift from 30°C to 37°C for 10 minutes**

Analysis of the upregulated genes by serotype 1 showed no enriched pathways based on the statistical threshold used (FDR <0.05). For upregulated genes by serotype 6B, four largely metabolic bacterial pathways were significantly enriched at 10 minutes in response to elevated temperature (**Figure 7.9A**). The most significantly upregulated pathways were for genes involved in pentose phosphate metabolism and PTS sugar transport. A high number of genes (n = 60) involved in central pneumococcal metabolism were also enriched (**Figure 7.9A**). The most significantly downregulated pathway for serotype 6B was pyruvate metabolism while serotype 1 significantly downregulated two-components gene systems (**Figure 7.9B**). Both strains downregulated oxocarboxylic acid metabolism pathway genes in response to temperature change from 30°C to 37°C.

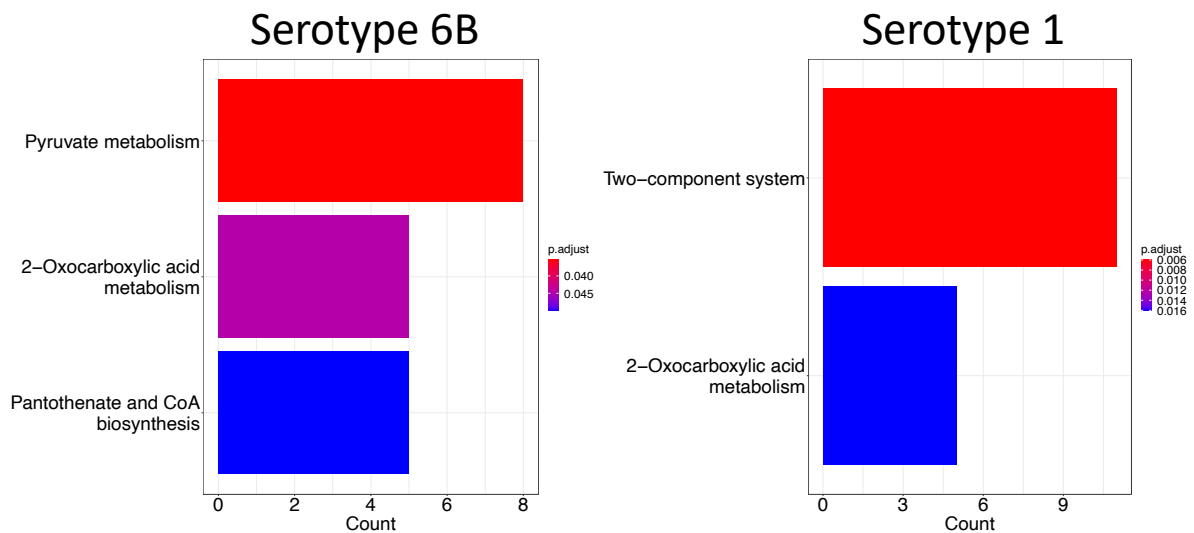
# Upregulated Pathways at 10 mins



**Figure 7.9A: Significantly upregulated pathways by *S. pneumoniae* serotype 6B (BHN418) following temperature change from 30°C to 37°C for 10 minutes.** Orthologs present in KEGG annotated strain for 670-6B (serotype 6B) were used. Pathway enrichment was performed by over representation analysis and pathways with FDR <0.05 were considered significantly enriched. Only pathways with FDR <0.05 were shown. Length of the bar corresponds to the number of DEGs in respective pathways and the colour indicates significance level with red being most statistically significant and blue being least significant.



## Downregulated Pathways: 10 mins



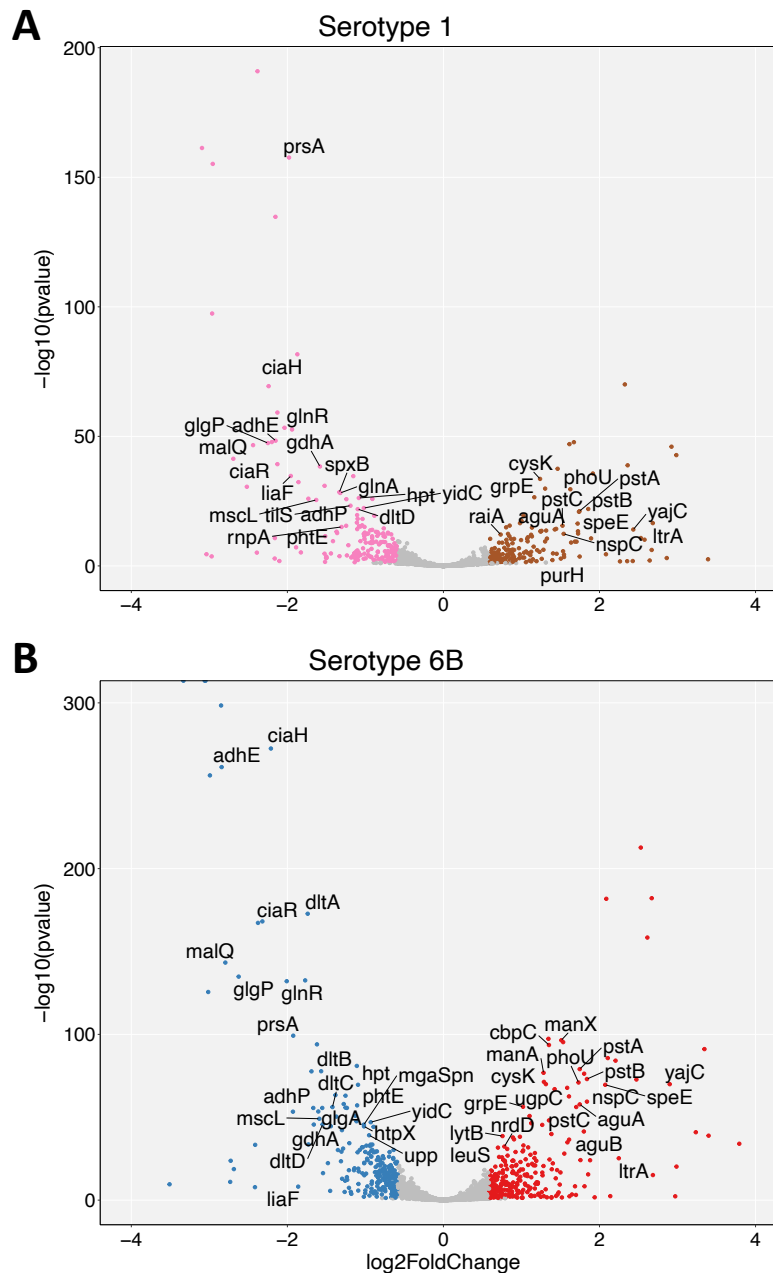
**Figure 7.9B: Significantly downregulated pathways by *S. pneumoniae* serotype 6B (BHN418) and serotype 1 (BVJ1JL) following temperature change from 30°C to 37°C for 10 minutes.** Orthologs present in KEGG annotated strain for 670-6B (serotype 6B) and INV104 (serotype 1) were used. Pathway enrichment was performed by over representation analysis and pathways with FDR <0.05 were considered significantly enriched. Only pathways with FDR <0.05 are shown. Length of the bar corresponds to the number of DEGs in respective pathways and colour indicates significance level with red being most statistically significant and blue being least significant within each plot.

### 7.3.8 Transcriptomics response of *S. pneumoniae* during temperature shift from 30°C to 37°C for 20 minutes

After 20 minutes incubation at elevated temperature of 37°C, the total number of differentially regulated genes in serotype 6B increased by ~11% (from 451 to 508) compared to the 10 minutes timepoint. The number of both upregulated and downregulated genes was increased in serotype 6B (**Figure 7.10A** and **Figure 7.10B**). In serotype 1, the total number of differentially regulated genes was similar at 10 minutes and 20 minutes with more genes being downregulated than upregulated at both timepoints. Similar to the results for the 10 minutes timepoint, about half (45.8%) of the total number of differentially expressed genes were commonly regulated in both strains.

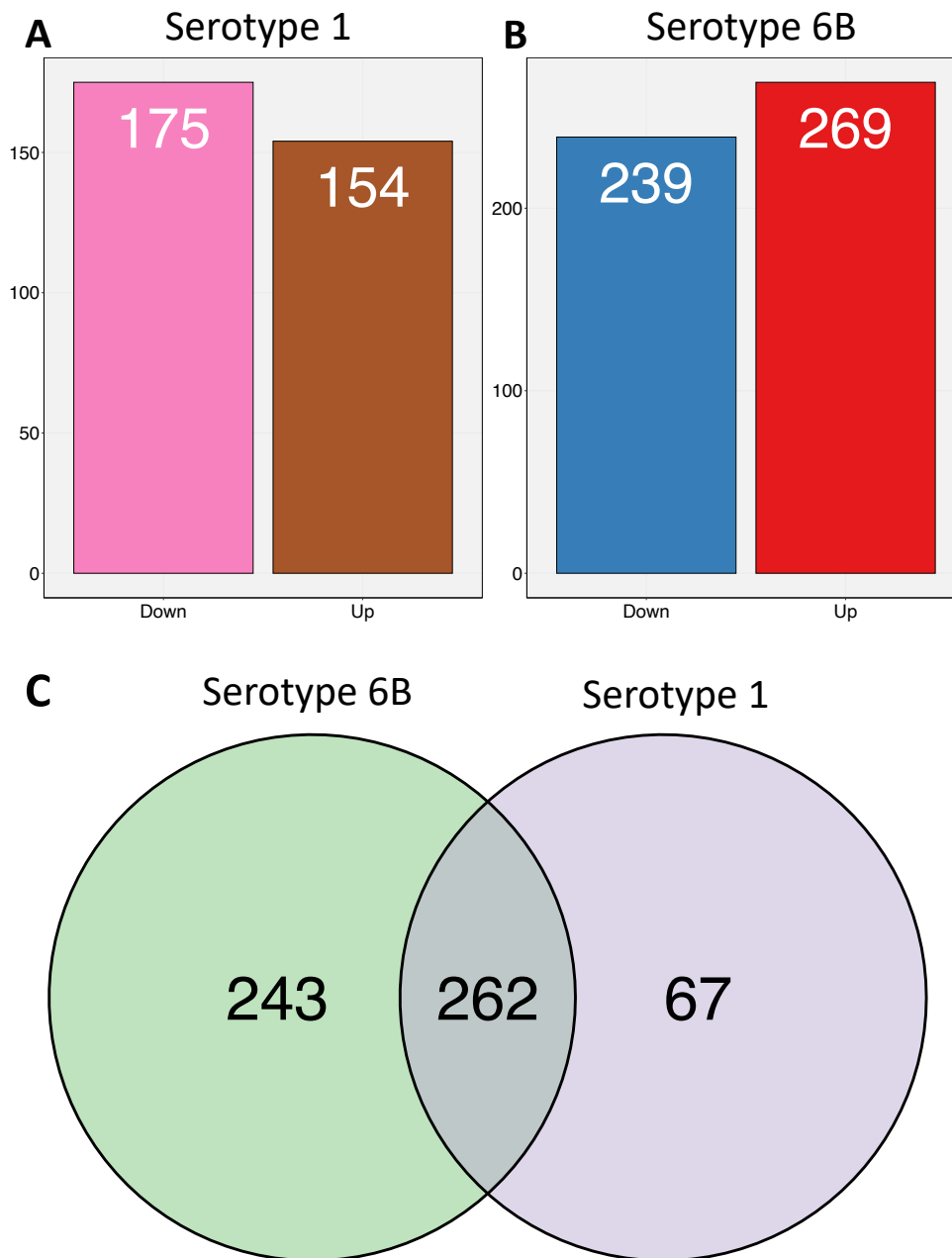
Serotype 6B increased the number of uniquely regulated genes by ~21 % (from 192 to 243) at 20 minutes compared to the 10 minutes timepoint (**Figure 7.10A** and **Figure-7.10B**). Genes that were upregulated at 20 minutes but not at 10 minutes for serotype 6B included the competence operon *comCDE*, kinases (e.g., *arcC*, *proB*, *galK*, *lacC*, *lysC*), several transporters (e.g., *JN057\_RS08815*, *JN057\_RS08510*, *lacA*, *JN057\_RS04075*), protein synthesis genes (e.g., *nrdD*, *rpmD*, *rfaA*, *rfaB*, *prmA*, *lepA*), hypothetical proteins (e.g., *JN057\_RS09655*, *JN057\_RS07510*), a phosphoglycerate mutase (*gpmA*), a phosphopentomutase (*deoB*), transcriptional regulators (e.g., *JN057\_RS06980*, *JN057\_RS05515*), peptidase T (*pepT*), and dehydrogenases (e.g., *pyrK*, *pyrDb*). Despite serotype 1 having similar number of differentially regulated genes at 20 minutes compared to 10 minutes, 63 genes were differentially regulated at 20 minutes but not at 10 minutes. Forty six percent (29/63) of these were upregulated. Uniquely upregulated genes at 20 minutes for serotype 1 Included the surface protein *cbpA*, ABC transporters (*manM*, *JN057\_RS01775*, *proWX*, *proV*), hypothetical proteins (e.g., *JN057\_RS09655*, *JN057\_RS00625*), transketolase (*tkt*), protein synthesis genes (*rpsN*, *rmpD*, *raiA*), dehydrogenases (e.g., *pyrK*). The most significantly upregulated genes by serotype 1 after 20 mins are listed on **Tables 7.7A** and **Table 7.7B**. Similar to growth at 10 minutes *htrA*, *ciaR*, *ciaH*, *aliA*, *glnR*, and *adhR* were among the most significantly downregulated genes by both strains at 20 minutes. *htrA* expression was downregulated by a further 23.2% (FC from -6.4 to -8.33) in serotype 6B compared to 10 minutes while serotype 1 was downregulated further by 6.3% (FC from -7.27 to -7.76) (**Tables 7.5B** and **Table 7.5C**). The *ciaRH* two-components genes fold changes at 20 minutes were similar to 10 minutes but more statistically significant (**Tables 7.7B** and **Table 7.7C**).

# DGE at 20 mins



**Figure 7.10A** Volcano plots showing differential gene expression (DGE) by *S. pneumoniae* serotype 1 (BVJ1JL) and serotype 6B (BHN418) following temperature shift from 30°C to 37°C for 20 minutes. Each point on the plot represents a gene with the up- (brown = serotype 1; red = serotype 6B) and down-regulated (pink = serotype 1; blue = serotype 6B) genes indicated by colour. Where available in the original annotation file, short gene names are annotated on volcano plots. DEGs were defined by fold change >1.5 and false discovery rate FDR <0.05.

# DGE at 20 mins



**Figure 7.10B: Total number of differentially expressed genes by of BVJ1JL (serotype 1) and BHN418UCL (serotype 6B) following temperature shift from 30°C to 37°C for 20 minutes.** Bar plots show the total number of up- (Brown = serotype 1; red = serotype 6B) and down- regulated (pink = serotype 1; blue = serotype 6B) genes. The Venn diagram shows the total number of genes that were shared or uniquely regulated by serotype 1 and serotype 6B. DEGs were defined by fold change >1.5 and false discovery rate FDR <0.05.

**Table 7.7A:** Top 40 differentially upregulated genes by *S. pneumoniae* serotype 6B (BHN418) following temperature shift from 30°C to 37°C for 20 minutes. Genes were ranked by statistical significance (P-adj). Top genes in both strains are in bold. RefSeq locus tags of the published sequence of BVJ1JL are shown (Betts *et al.* 2021).

locus	Gene	FC	P-adj	Annotation
<b>Serotype 6B (BHN418)</b>				
<b>JN057_RS10280</b>	<b>pstS1</b>	5.77	2.01E-213	Phosphate ABC transporter periplasmic phosphate-binding protein PstS1
<b>JN057_RS05700</b>	<b>gapN</b>	6.36	6.93E-183	NADP-dependent glyceraldehyde-3-phosphate dehydrogenase
JN057_RS07405	aliB	4.25	1.62E-182	peptide ABC transporter substrate-binding protein
<b>JN057_RS01840</b>	<b>clpL</b>	6.12	3.86E-159	ATP-dependent Clp protease ATP-binding subunit
<b>JN057_RS02990</b>		2.54	4.67E-98	ferric reductase-like transmembrane domain-containing protein
<b>JN057_RS01605</b>	<b>manX</b>	2.85	3.56E-97	PTS sugar transporter subunit IIB
<b>JN057_RS10060</b>	<b>adh</b>	2.90	4.64E-96	zinc-dependent alcohol dehydrogenase family protein
<b>JN057_RS02040</b>	<b>cbpC</b>	2.55	2.63E-94	choline-binding protein CbpC
JN057_RS01780		10.16	7.51E-92	PTS system mannose/fructose/sorbose family transporter subunit IID
<b>JN057_RS04390</b>	<b>cad</b>	4.31	1.98E-86	aminotransferase class I/II-fold pyridoxal phosphate-dependent enzyme
<b>JN057_RS04405</b>	<b>lys1</b>	4.61	7.08E-85	saccharopine dehydrogenase family protein
<b>JN057_RS10290</b>	<b>pstA</b>	3.36	8.87E-80	phosphate ABC transporter permease PstA
<b>JN057_RS03575</b>	<b>manA</b>	2.43	1.51E-77	mannose-6-phosphate isomerase, class I
JN057_RS08320	scrA	3.49	5.30E-77	sucrose-specific PTS transporter subunit IIBC
<b>JN057_RS10295</b>	<b>pstB</b>	3.58	8.84E-74	phosphate ABC transporter ATP-binding protein PstB

(Table Continued)

# Chapter 7: Transcriptomic response to temperature change

<i>JN057_RS01740</i>	<i>eda</i>	5.55	1.69E-73	bifunctional 4-hydroxy-2-oxoglutarate aldolase/2-dehydro-3-deoxy-phosphogluconate aldolase
<i>JN057_RS10875</i>	<i>cysK</i>	2.44	4.75E-72	cysteine synthase A
<i>JN057_RS10300</i>	<i>phoU</i>	3.32	1.06E-71	phosphate signalling complex protein PhoU
<i>JN057_RS07495</i>		2.49	1.06E-70	NAD(P)H-dependent oxidoreductase
<i>JN057_RS01785</i>	<i>yajC</i>	7.47	1.11E-70	preprotein translocase subunit YajC
<i>JN057_RS04400</i>	<i>speE</i>	4.21	2.63E-70	polyamine aminopropyltransferase
<i>JN057_RS07660</i>	<i>ugpC</i>	3.00	1.77E-68	sn-glycerol-3-phosphate ABC transporter ATP-binding protein UgpC
<i>JN057_RS08330</i>	<i>scrR</i>	2.69	1.12E-67	LacI family DNA-binding transcriptional regulator
<i>JN057_RS08325</i>	<i>scrB</i>	3.05	2.61E-63	glycoside hydrolase family 32 protein
<i>JN057_RS04410</i>	<i>nspC</i>	3.58	3.77E-60	carboxynorspermidine decarboxylase
<i>JN057_RS04415</i>	<i>aguA</i>	3.36	1.67E-58	agmatine deiminase
<i>JN057_RS02615</i>	<i>grpE</i>	2.03	4.83E-57	nucleotide exchange factor GrpE
<i>JN057_RS10285</i>	<i>pstC</i>	3.25	6.74E-57	phosphate ABC transporter permease subunit PstC
<i>JN057_RS01600</i>	<i>manM</i>	2.15	1.78E-51	PTS mannose/fructose/sorbose transporter subunit IIC
<i>JN057_RS02470</i>	<i>mtsA</i>	2.55	7.52E-49	ECF-type riboflavin transporter substrate-binding protein
<i>JN057_RS03055</i>		2.18	4.05E-47	amino acid ABC transporter substrate-binding protein
<i>JN057_RS04065</i>		2.41	4.09E-46	BMP family protein
<i>JN057_RS04420</i>	<i>aguB</i>	3.48	3.94E-42	N-carbamoylputrescine amidase
<i>JN057_RS01775</i>		9.41	1.07E-41	PTS mannose/fructose/sorbose/N-acetyl galactosamine transporter subunit IIC
<i>JN057_RS01040</i>	<i>metN</i>	2.61	1.32E-40	methionine ABC transporter ATP-binding protein
<i>JN057_RS01765</i>	<i>ugl</i>	10.54	1.30E-39	glycoside hydrolase family 88 protein

*(Table Continued)*

<i>JN057_RS04625</i>	<i>lytB</i>	1.69	2.27E-39	glucosaminidase domain-containing protein
<b><i>JN057_RS02170</i></b>		1.97	5.95E-39	carboxymuconolactone decarboxylase family protein
<i>group_614</i>		1.85	1.57E-38	hypothetical protein
<b><i>JN057_RS03655</i></b>		1.87	2.19E-37	CBS domain-containing protein

---

**Table 7.7B:** Top 40 differentially upregulated genes by *S. pneumoniae* serotype 1 (BVJ1JL) following temperature shift from 30°C to 37°C for 20 minutes. Genes were ranked by statistical significance (P-adj). Top genes in both strains are in bold. RefSeq locus tags of the published sequence of BVJ1JL are shown (Betts *et al.* 2021).

locus	Gene	FC	P-adj	Annotation
<b>Serotype 1 (BVJ1JL)</b>				
<b>JN057_RS10280</b>	<b>pstS1</b>	5.01	8.89E-71	Phosphate ABC transporter periplasmic phosphate-binding protein PstS1
<b>JN057_RS07495</b>		3.19	1.78E-48	NAD(P)H-dependent oxidoreductase
<i>group_324</i>		3.06	9.95E-48	amino acid ABC transporter ATP-binding protein
<b>JN057_RS01840</b>	<b>clpL</b>	7.58	9.71E-47	ATP-dependent Clp protease ATP-binding subunit
<b>JN057_RS05700</b>	<b>gapN</b>	7.92	1.74E-43	NADP-dependent glyceraldehyde-3-phosphate dehydrogenase
<i>group_1371</i>		5.14	1.49E-39	unannotated protein
<i>group_1478</i>		2.76	3.25E-38	unannotated protein
<b>JN057_RS10060</b>	<b>adh</b>	3.77	2.17E-36	zinc-dependent alcohol dehydrogenase family protein
<b>JN057_RS10875</b>	<b>cysK</b>	2.36	2.63E-34	cysteine synthase A
<b>JN057_RS04065</b>		2.47	1.46E-30	BMP family protein
<b>JN057_RS10300</b>	<b>phoU</b>	3.09	2.65E-30	phosphate signalling complex protein PhoU
<b>JN057_RS02615</b>	<b>grpE</b>	2.24	3.22E-27	nucleotide exchange factor GrpE
<b>JN057_RS10295</b>	<b>pstB</b>	3.62	1.06E-22	phosphate ABC transporter ATP-binding protein PstB
<b>JN057_RS10290</b>	<b>pstA</b>	3.35	7.75E-22	phosphate ABC transporter permease PstA
<i>JN057_RS09645</i>		3.33	1.34E-21	hypothetical protein
<i>JN057_RS03830</i>		2.04	3.96E-20	membrane protein
<i>JN057_RS08615</i>		2.00	4.34E-18	hypothetical protein
<i>JN057_RS01845</i>	<i>ltrA</i>	6.42	3.01E-17	group II intron reverse transcriptase/maturase
<i>JN057_RS05910</i>		1.97	3.20E-17	phosphocarrier protein <i>HPr</i>



(Table Continued)

<b>JN057_RS04390</b>	<b><i>cad</i></b>	3.29	4.61E-17	aminotransferase class I/II-fold pyridoxal phosphate-dependent enzyme
<b>JN057_RS03055</b>		1.80	2.80E-16	amino acid ABC transporter substrate-binding protein
<b>JN057_RS10285</b>	<b><i>pstC</i></b>	2.88	2.96E-16	phosphate ABC transporter permease subunit PstC
<b>JN057_RS02170</b>		1.75	1.30E-15	carboxymuconolactone decarboxylase family protein
<b>JN057_RS02040</b>	<b><i>cbpC</i></b>	2.20	1.64E-15	choline-binding protein CbpC
<b>JN057_RS04415</b>	<b><i>aguA</i></b>	2.72	5.10E-15	agmatine deiminase
<b>JN057_RS02990</b>		2.69	8.53E-15	ferric reductase-like transmembrane domain-containing protein
<b>JN057_RS01785</b>	<b><i>yajC</i></b>	5.40	8.96E-15	preprotein translocase subunit YajC
<b>JN057_RS04420</b>	<b><i>aguB</i></b>	2.49	2.28E-14	N-carbamoylputrescine amidase
<b>JN057_RS07660</b>	<b><i>ugpC</i></b>	2.39	2.96E-14	sn-glycerol-3-phosphate ABC transporter ATP-binding protein UgpC
<b>JN057_RS10335</b>		2.37	3.99E-14	rhomboid family intramembrane serine protease
<b>JN057_RS04400</b>	<b><i>speE</i></b>	3.30	4.08E-14	polyamine aminopropyltransferase
<b>JN057_RS01605</b>	<b><i>manX</i></b>	2.37	5.65E-14	PTS sugar transporter subunit IIB
<b>JN057_RS04405</b>	<b><i>Lys1</i></b>	3.31	3.33E-13	saccharopine dehydrogenase family protein
<b>JN057_RS04410</b>	<b><i>nspC</i></b>	2.91	4.26E-13	carboxynorspermidine decarboxylase
<b>JN057_RS04835</b>	<b><i>pepT</i></b>	1.78	4.30E-13	peptidase T
<b>JN057_RS10855</b>	<b><i>raiA</i></b>	1.66	7.82E-13	ribosome-associated translation inhibitor RaiA
<b>JN057_RS01530</b>		2.18	9.54E-13	LLM class flavin-dependent oxidoreductase
<b>JN057_RS01050</b>		1.80	3.58E-12	MptD family putative ECF transporter S component
<b>JN057_RS03655</b>		2.19	4.12E-12	CBS domain-containing protein
<b>JN057_RS03575</b>	<b><i>manA</i></b>	2.12	8.95E-12	mannose-6-phosphate isomerase, class I

**Table 7.7C:** Top 20 differentially downregulated genes by *S. pneumoniae* serotype 6B (BHN418) following temperature shift from 30°C to 37°C for 10 minutes. Genes were ranked by statistical significance (P-adj). Top genes in both strains are in bold. RefSeq locus tags of the published sequence of BVJ1JL are shown (Betts *et al.* 2021).

locus	Gene	FC	padj	Annotation
<b>Serotype 6B (BHN418)</b>				
<b>JN057_RS04930</b>		-10.07	0.00E+00	DUF1002 domain-containing protein
<b>JN057_RS11045</b>	<b>htrA</b>	-8.33	0.00E+00	trypsin-like peptidase domain-containing protein
<b>JN057_RS03040</b>		-8.28	0.00E+00	M57 family metalloprotease
<b>JN057_RS08285</b>		-7.21	3.80E-299	membrane protein
<b>JN057_RS03875</b>	<b>ciaH</b>	-4.63	3.56E-273	two-component system sensor histidine kinase
<b>JN057_RS09915</b>	<b>adhE</b>	-7.17	5.52E-262	bifunctional acetaldehyde-CoA/alcohol dehydrogenase
JN057_RS08280		-7.96	5.64E-257	ABC transporter ATP-binding protein
JN057_RS10725	<i>dltA</i>	-3.34	1.66E-173	D-alanine--poly(phosphoribitol) ligase subunit
<b>JN057_RS03870</b>	<b>ciaR</b>	-4.99	6.94E-169	two-component system response regulator
<b>JN057_RS11050</b>		-5.19	5.53E-168	ParB/RepB/Spo0J family partition protein
<b>JN057_RS10395</b>	<b>malQ</b>	-6.94	4.83E-144	4-alpha-glucanotransferase
<b>JN057_RS10390</b>	<b>glgP</b>	-6.17	1.44E-135	glycogen/starch/alpha-glucan family phosphorylase
JN057_RS06185	<i>glnHP5</i>	-3.41	2.65E-133	ABC transporter substrate-binding protein/permease
<b>JN057_RS02545</b>	<b>glnR</b>	-4.02	8.40E-133	transcriptional repressor GlnR
JN057_RS08275	<i>gntR</i>	-8.08	2.69E-126	GntR family transcriptional regulator
<b>JN057_RS04705</b>	<b>prsA</b>	-3.80	6.16E-100	peptidylprolyl isomerase PrsA
<b>JN057_RS01080</b>		-3.07	1.04E-94	divalent metal cation transporter
JN057_RS04800	<i>yfnA</i>	-2.16	1.06E-81	amino acid permease
JN057_RS10720	<i>dltB</i>	-2.96	1.71E-78	D-alanyl-lipoteichoic acid biosynthesis protein
JN057_RS01185	<i>folC</i>	-3.23	2.07E-78	Bifunctional dihydrofolate synthase

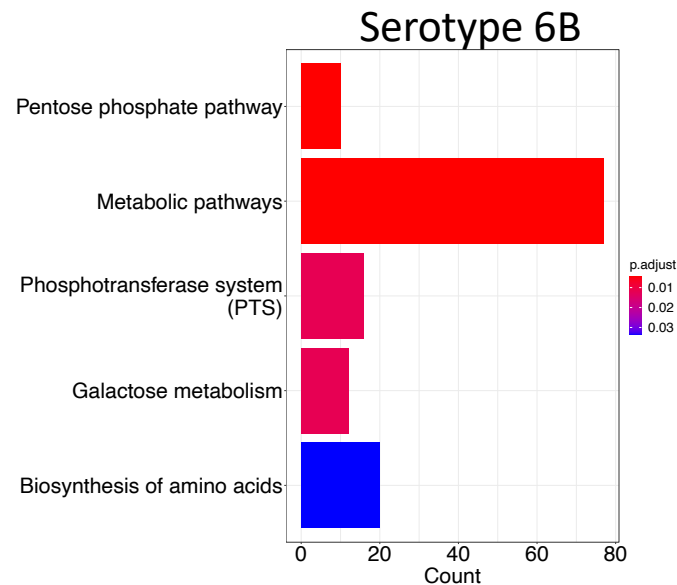
**Table 7.7D:** Top 20 differentially downregulated genes by *S. pneumoniae* serotype 1 (BVJ1JL) following temperature shift from 30°C to 37°C for 20 minutes. Genes were ranked by statistical significance (padj). Top genes in both strains are in bold. RefSeq locus tags of the published sequence of BVJ1JL are shown (Betts *et al.* 2021).

locus	Gene	FC	padj	Annotation
<b>Serotype 1 (BVJ1JL)</b>				
<b>JN057_RS11050</b>		-5.22	1.21E-191	ParB/RepB/Spo0J family partition protein
<b>JN057_RS04930</b>		-8.54	5.00E-162	DUF1002 domain-containing protein
<b>JN057_RS04705</b>	<b>prsA</b>	-3.94	2.43E-158	peptidylprolyl isomerase PrsA
<b>JN057_RS11045</b>	<b>htrA</b>	-7.76	6.88E-156	trypsin-like peptidase domain-containing protein
JN057_RS04360		-4.45	1.84E-135	PspC domain-containing protein
<b>JN057_RS03040</b>		-7.80	3.89E-98	M57 family metalloprotease
<b>JN057_RS03875</b>	<b>ciaH</b>	-3.66	2.15E-82	two-component system sensor histidine kinase
<b>JN057_RS08285</b>		-4.72	4.27E-70	membrane protein
JN057_RS04230		-4.37	6.42E-60	hypothetical protein
JN057_RS01980	<i>aliA</i>	-4.10	4.77E-54	peptide ABC transporter substrate-binding protein
<b>JN057_RS02545</b>	<b>glnR</b>	-3.84	2.25E-53	transcriptional repressor GlnR
<b>JN057_RS09915</b>	<b>adhE</b>	-4.44	4.64E-49	bifunctional acetaldehyde-CoA/alcohol dehydrogenase
<b>JN057_RS01080</b>		-4.59	1.56E-48	divalent metal cation transporter
<b>JN057_RS10390</b>	<b>glgP</b>	-4.76	3.83E-48	glycogen/starch/alpha-glucan family phosphorylase
<b>JN057_RS10395</b>	<b>malQ</b>	-5.42	2.46E-47	4-alpha-glucanotransferase
JN057_RS08275	<i>gntR</i>	-6.47	4.42E-42	GntR family transcriptional regulator
<b>JN057_RS03870</b>	<b>ciaR</b>	-4.37	5.34E-40	two-component system response regulator
JN057_RS06490	<i>gdhA</i>	-3.00	4.25E-39	NADP-specific glutamate dehydrogenase
JN057_RS02085	<i>liaF</i>	-3.88	1.91E-35	cell wall-active antibiotics response protein
JN057_RS04610		-2.23	2.25E-35	VOC family protein

### **7.3.9 Biological pathways enriched by *S. pneumoniae* during temperature shift from 30°C to 37°C for 20 minutes**

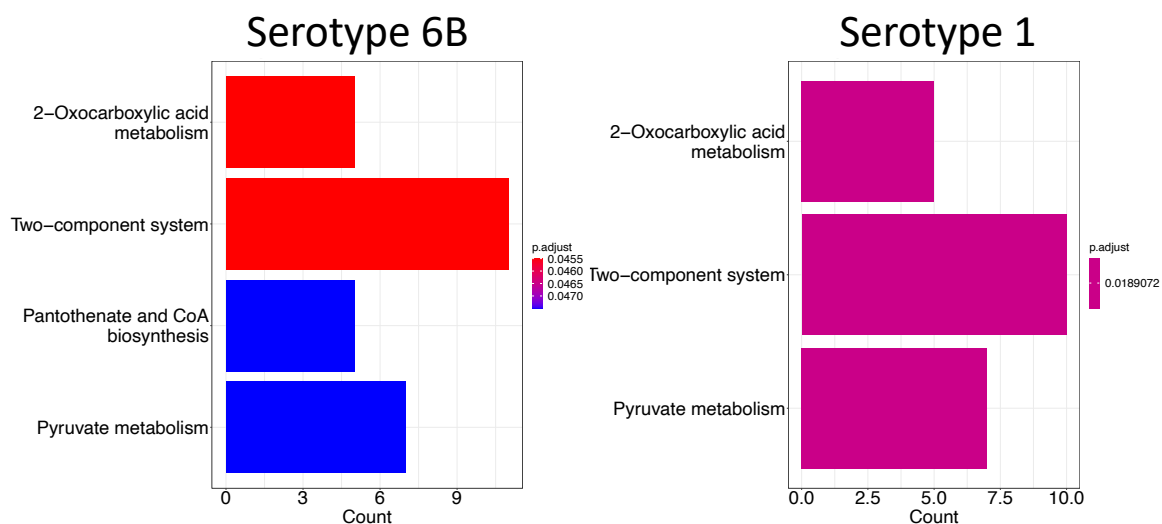
Similar to the 10 minutes incubation, genes upregulated by serotype 1 were not statistically enriched for any KEGG biological pathways. For serotype 6B, enrichment of metabolic pathways was more significantly enriched with total gene counts in this pathway increasing from 60 at 10 minutes to 77 at 20 minutes (**Figure 7.11A**). In addition to pentose phosphate pathway and PTS pathways, which were also upregulated pathways at 10 minutes, galactose metabolism and amino acid synthesis pathways were upregulated by serotype 6B at 20 minutes. Downregulation of competence gene pathways was statistically significant for both strains after 20 minutes in contrast to 10 minutes when only serotype 1 achieved statistical significance for this pathway. Similar to 10 minutes, Oxocarboxylic acid and pyruvate metabolism pathways remained significantly downregulated by both serotype 1 and serotype 6B after 20 minutes (**Figure 7.11B**).

# Upregulated Pathways at 20 mins



**Figure 7.11A: Significantly upregulated pathways by *S. pneumoniae* serotype 6B (BHN418) following temperature change from 30°C to 37°C for 20 minutes.** Orthologs present in KEGG annotated strain for 670-6B (serotype 6B) were used. Pathway enrichment was performed by over representation analysis and pathways with FDR <0.05 were considered significantly enriched. Only pathways with FDR <0.05 are shown. Length of the bar corresponds to the number of DEGs in respective pathways and colour indicates significance level with red being most statistically significant and blue being least significant.

## Downregulated Pathways: 20 mins

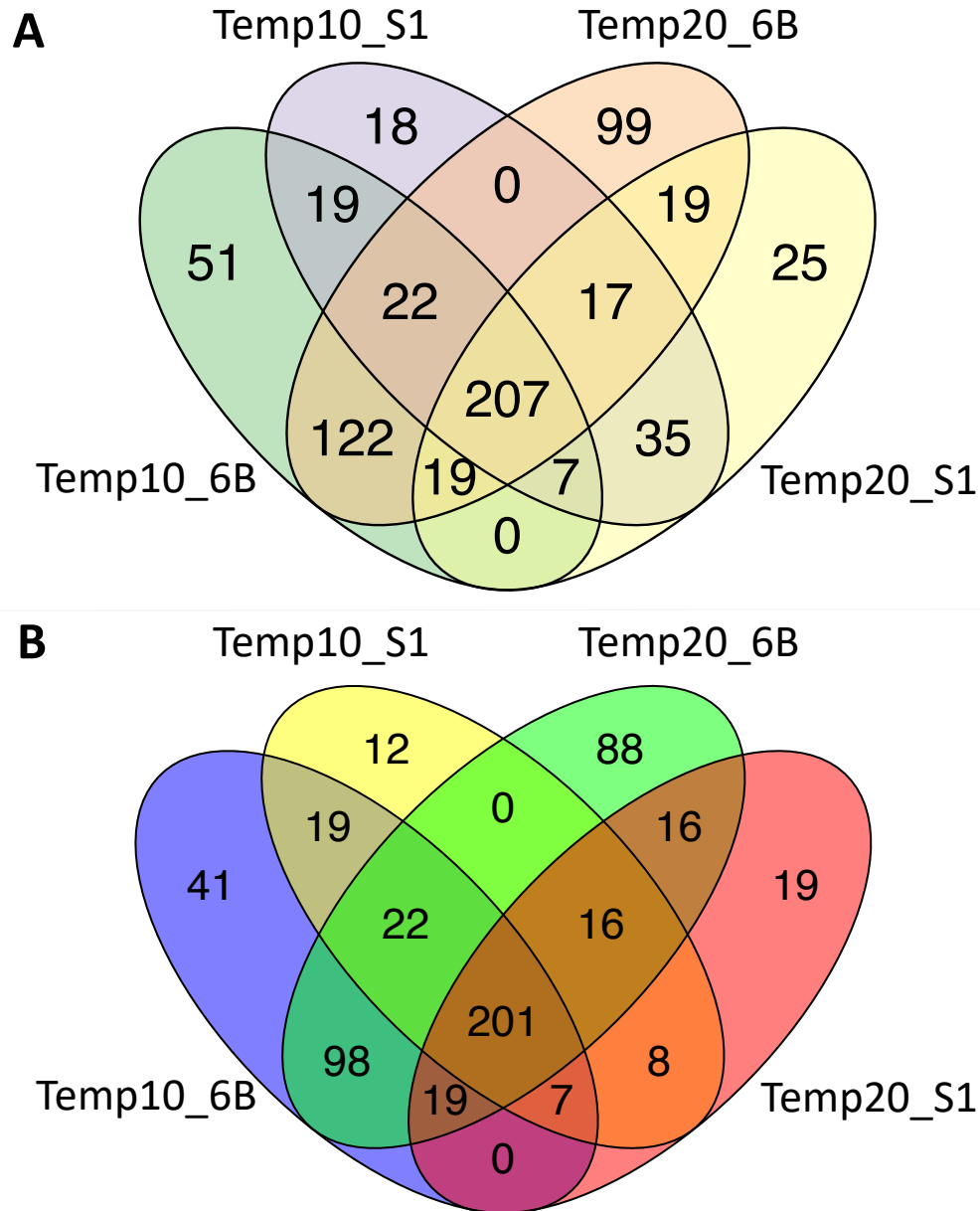


**Figure 7.11B: Significantly downregulated pathways by *S. pneumoniae* serotype 6B (BHN418) and serotype 1 (BVJ1JL) following temperature change from 30°C to 37°C for 20 minutes.** Orthologs present in KEGG annotated strain for 670-6B (serotype 6B) and INV104 (serotype 1) were used. Pathway enrichment was performed by over representation analysis and pathways with FDR <0.05 were considered significantly enriched. Only pathways with FDR <0.05 are shown. Length of the bar corresponds to the number of DEGs in respective pathways and colour indicates significance level with red being most statistically significant and within each plot.

### 7.3.10 Comparison of transcriptomes of *S. pneumoniae* strains at all timepoints following growth at elevated temperature

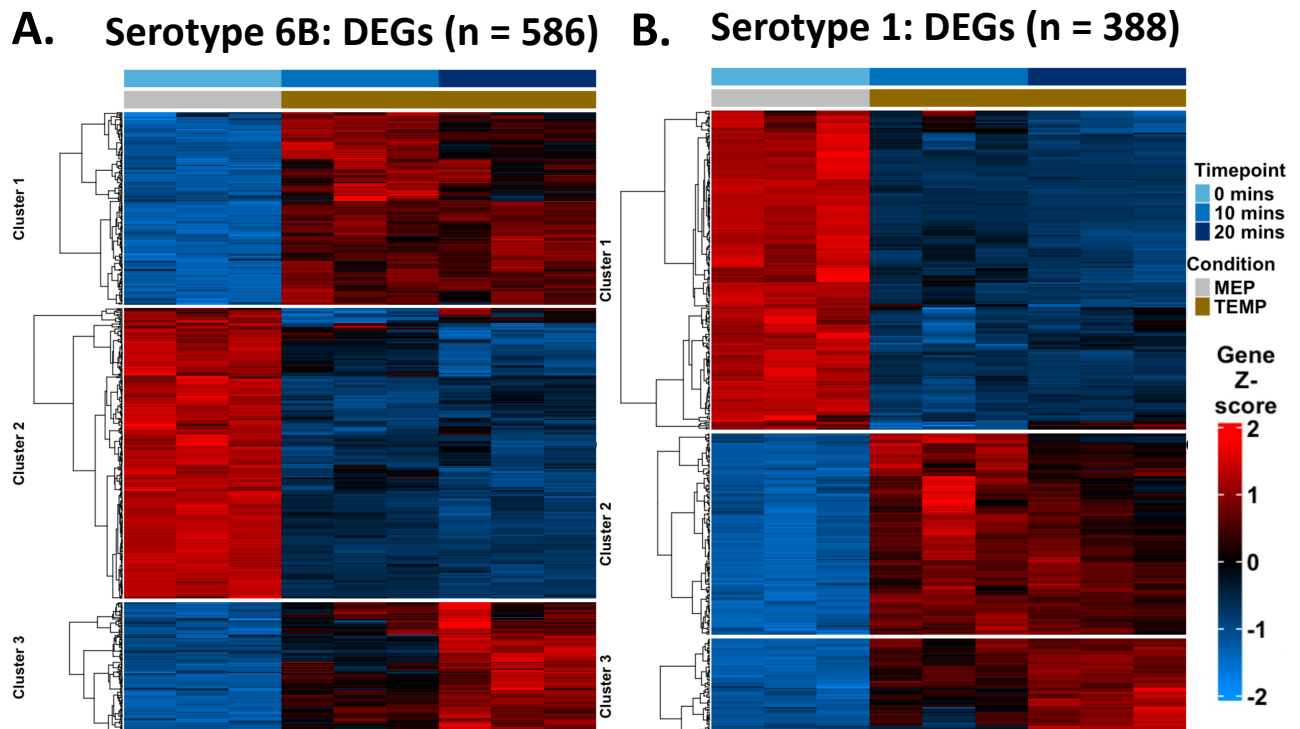
**Figure 7.12A-A** shows the association in total number of differentially expressed genes in all conditions for serotype 1 and serotype 6B. Restricting this analysis to the core genes that were present in both strains in **Figure 7.12A-B** recapitulated the same pattern as before albeit with a lower number of DEGs. These results suggests that the differences in the phenotypic adaptation of serotype 1 and serotype 6B, including growth difference, may be due to nuanced differences in how the strains specifically fine-tune the expression of core genes upon exposure to environment stimuli. A total number of 586 and 388 genes were differentially expressed in at least one of our studied conditions for serotype 6B and serotype 1 respectively. Unsupervised hierarchical clustering identified three main clusters: a cluster of genes upregulated at both 10 minutes and 20 minutes, a cluster of genes downregulated at both 10 minutes and 20 minutes; and a cluster where genes were upregulated at 10 minutes but upregulated further at 20 minutes (**Figure 7.12B**).

# Differential expression of orthologs



**Figure 7.12A:** Venn diagram of differentially expressed genes at all conditions for *S. pneumoniae* serotype 6B (BHN418) and serotype 1 (BVJ1JL) following temperature change from 30°C to 37°C. **A).** The total number of DEGs at each condition using the entire gene sets of each strain. **B).** Number of DEG with only orthologs present in both strains (core genes only).





**Figure 7.12B: Unsupervised hierarchical clustering analysis of DEGs of *S. pneumoniae* serotype 6B (BHN418) and serotype 1 (BVJ1JL) following temperature change from 30°C to 37°C for 10 and 20 minutes.** Heatmap based on z-scores of gene expression was constructed for all genes that were differentially expressed in at least one study condition (n = 586 for serotype 6B and n = 388 for serotype 1). Upper tracks and legends indicate triplicates samples of each condition.

### 7.3.11 Transcriptomics profiles of selected *S. pneumoniae* genes in response to temperature change

In their analysis of the transcriptome of D39 transcriptome, Aprianto and colleagues identified 19 temperature responsive genes that were upregulated by *S. pneumoniae* D39 strain at elevated temperatures (Aprianto *et al.* 2018). We corroborated these finding for some of the highly conserved pneumococcal proteins including *dnaK*, *hrcA*, *hrcA*, *grpE*, and *dnaJ* (**Figure 7.13A**). However, we also observed both time dependent and strain dependent differences in the expression of these genes in the clinical strains used here. For example, in contrast to their finding for the D39 strain, the glutamine ABC transporter *glnQ2* was downregulated at all timepoints for serotype 1 and serotype 6B after exposure to elevated temperatures of 37°C. Other temperature response genes such as *cstR*, *clpC*, *thiX*, and *clpP*, which were upregulated in the D39 transcriptomic study, were not differentially expressed by either *S. pneumoniae* serotype 1 or serotype 6B following transfer from 30°C to 37°C (**Figure 7.13A**). The gene expression profiles of temperature response genes in serotype 1 and serotype 6B strains are shown in **Figure 7.13A**. The expression profiles of a selection of virulence, competence, and oxidative stress genes is illustrated in **Figure 7.13B**.

To further demonstrate that transcriptomic experiments captured adaptation to specific environmental exposures that were being investigated, and not just general *S. pneumoniae* stress responses, we profiled the expression of iron response genes (**Figure 7.13C**). In Chapter 6, it was shown that the expression of some of these iron response genes by *S. pneumoniae* was highly correlated to iron availability in growth media. Apart from *copA* which was upregulated by serotype 6B but not serotype 1, none of the iron response genes were upregulated during growth in elevated temperatures, highlighting redundancy of these iron-response genes in high temperature environments and underscoring the highly coordinated and specific transcriptomic adaptation by *S. pneumoniae* in response to environmental exposure. The *copA* gene is one of a three-gene *cop* operon which has been suggested as essential for *S. pneumoniae* homeostasis in high copper environments and also contribute to pneumococcal virulence (Shafeeq *et al.* 2011). Strains that lack *copA* showed attenuated virulence in murine models of pneumococcal pneumonia (Shafeeq *et al.* 2011). Our data suggests that *copA* may also

play a role in *S. pneumoniae* adaptation to elevated temperatures in a strain dependent manner (in serotype 6B but not serotype 1) (**Figure 7.13C**).

Transcriptomic profile of “temperature responsive” genes in clinical strains of serotype 1 and serotype 6B. All genes previously found to be upregulated in D39 strain (Aprianto *et al.*, 2018)

Gene	Locus tag (BVJ1JL)	Serotype 6B (BHN418UCL)		Serotype 1 (BVJ1JL)	
		Temp10	Temp20	Temp10	Temp20
<i>groL</i>	JN057_RS09315	↑	↑	↑	↔
<i>groES</i>	JN057_RS09320	↑	↔	↑	↔
<i>clpP</i>	JN057_RS03620	↔	↔	↔	↔
<i>clpL</i>	JN057_RS01840	↑	↑	↑	↑
<i>dnaK</i>	JN057_RS02620	↑	↑	↑	↑
<i>hrcA</i>	JN057_RS02610	↑	↑	↑	↑
<i>grpE</i>	JN057_RS02615	↑	↑	↑	↑
<i>dnaJ</i>	JN057_RS02630	↑	↑	↑	↑
<i>clpC</i>	JN057_RS10800	↔	↔	↔	↔
<i>ctsR</i>	JN057_RS10805	↔	↔	↔	↔
<i>glnQ2</i>	JN057_RS03010	↓	↓	↓	↓
<i>hk06</i>	JN057_RS10790	↑	↔	↔	↔
<i>rr06</i>	JN057_RS10795	↑	↑	↑	↔
<i>thiX</i>	JN057_RS10820	↔	↔	↔	↔
<i>thiY</i>	JN057_RS10815	↑	↑	↑	↑
<i>thiZ</i>	JN057_RS10810	↑	↔	↑	↑
<i>SPV_2027</i>	JN057_RS10825	↔	↔	↑	↔

Upregulated = ↑ Not Differentially Expressed = ↔ Downregulated = ↓

Figure 7.13A: Transcriptomics profile of previously identified *S. pneumoniae* temperature response genes in BVJ1JL (serotype 1) and BHN418 (serotype 6B) strains at 10 and 20 minutes in elevated temperatures. Heat response genes were Identified by RNA-Seq using the D39 strain (serotype 2) (Aprianto *et al.* 2018). Differentially expressed genes (DEGs) were defined as genes with fold-change >1.5 and false discovery rate (FDR) <0.05.

VIRULENCE GENES					
Gene	Locus tag (BVJ1JL)	Serotype 6B (BHN418UCL)		Serotype 1 (BVJ1JL)	
		Temp10	Temp20	Temp10	Temp20
<i>psaA</i>	JN057_RS07965	↔	↔	↔	↔
<i>lytA</i>	JN057_RS09455	↔	↔	↔	↔
<i>ply</i>	JN057_RS09390	↔	↔	↔	↔
<i>cpsA</i>	JN057_RS01885	↔	↔	↔	↔
<i>nanA</i>	JN057_RS08170	↔	↔	↔	↔
<i>nanB</i>	JN057_RS08135	↔	↔	↔	↔
<i>aliA</i>	JN057_RS01980	↓	↓	↓	↓
<i>aliB</i>	JN057_RS07405	↑	↑	↑	↑

COMPETENCE GENES					
Gene	Locus tag (BVJ1JL)	Serotype 6B (BHN418UCL)		Serotype 1 (BVJ1JL)	
		Temp10	Temp20	Temp10	Temp20
<i>ciaR</i>	JN057_RS03870	↓	↓	↓	↓
<i>ciaH</i>	JN057_RS03875	↓	↓	↓	↓
<i>comA</i>	JN057_RS00510	↔	↔	↔	↔
<i>comB</i>	JN057_RS00515	↔	↔	↔	↔
<i>comE</i>	JN057_RS11020	↔	↑	↔	↔
<i>comD</i>	JN057_RS11025	↔	↑	↔	↔
<i>comC</i>	JN057_RS11030	↔	↑	↔	↔
<i>dprA</i>	JN057_RS06305	↔	↓	↔	↔

OXIDATIVE STRESS RESPONSE GENES					
Gene	Locus tag (BVJ1JL)	Serotype 6B (BHN418UCL)		Serotype 1 (BVJ1JL)	
		Temp10	Temp20	Temp10	Temp20
<i>spxB</i>	JN057_RS03555	↓	↓	↓	↓
<i>spx</i>	JN057_RS01160	↔	↓	↔	↓
<i>lctO</i>	JN057_RS03470	↔	↔	↔	↔
<i>Dpr</i>	JN057_RS07620	↓	↓	↔	↔
<i>adhE</i>	JN057_RS09915	↓	↓	↓	↓

Upregulated = ↑ Not Differentially Expressed = ↔ Downregulated = ↓

**Figure 7.13B: Transcriptomics profile of a cross-section of gene classes across *S. pneumoniae* serotype 1 (BVJ1JL) and serotype 6B (BHN418) at 10 and 20 minutes in elevated temperatures.** Differentially expressed genes (DEGs) were defined as genes with fold-change >1.5 and false discovery rate (FDR) <0.05.

## Transcriptomic profile transition metals genes in serotype 1 and serotype 6B

Gene	Locus tag (BVJ1JL)	Serotype 6B (BHN418UCL)		Serotype 1 (BVJ1JL)	
		Temp10	Temp20	Temp10	Temp20
<i>piuA</i>	JN057_RS09020	↔	↔	↔	↔
<i>piuB</i>	JN057_RS09005	↔	↔	↔	↔
<i>piuC</i>	JN057_RS09010	↔	↔	↔	↔
<i>piuD</i>	JN057_RS09015	↔	↔	↔	↔
<i>piaA</i>	JN057_RS05280	↔	↔	↔	↔
<i>piaB</i>	JN057_RS05285	↔	↔	↔	↔
<i>piaC</i>	JN057_RS05290	↔	↔	↔	↓
<i>piaD</i>	JN057_RS05295	↔	↔	↔	↔
<i>ritR</i>	JN057_RS02035	↔	↔	↔	↔
<i>psaA</i>	JN057_RS07965	↔	↔	↔	↔
<i>psaB</i>	JN057_RS07955	↔	↔	↔	↔
<i>psaC</i>	JN057_RS07960	↔	↔	↔	↔
<i>copY</i>	JN057_RS03540	↔	↔	↔	↔
<i>cupA</i>	JN057_RS03545	↔	↔	↔	↔
<i>copA</i>	JN057_RS03550	↑	↑	↔	↔
<i>adcA</i>	JN057_RS10690	↔	↔	↔	↔
<i>adcAll</i>	JN057_RS04805	↔	↔	↔	↔
<i>adcB</i>	JN057_RS10695	↔	↔	↔	↔
<i>adcC</i>	JN057_RS10700	↔	↔	↔	↔
<i>adcR</i>	JN057_RS10705	↔	↔	↔	↔
<i>czcD</i>	JN057_RS08950	↓	↔	↔	↔
<i>sodA</i>	JN057_RS03705	↔	↔	↔	↔
<i>mntE</i>	JN057_RS07530	↔	↔	↔	↔
<i>Gls24</i>	JN057_RS08665	↔	↔	↔	↔
<i>Dpr</i>	JN057_RS07620	↓	↓	↔	↔

Upregulated = ↑ Not Differentially Expressed = ↔ Downregulated = ↓

Figure 7.13C: Transcriptomics profile of previously identified transition metals response genes (Table 6-1) across *S. pneumoniae* serotype 1 (BVJ1JL) and serotype 6B (BHN418) strains at 10 and 20 minutes in elevated temperatures. Differentially expressed genes (DEGs) were defined as genes with fold-change >1.5 and false discovery rate (FDR) <0.05.

## 7.4 Discussion

The results presented in this chapter represent the first inter-strain comparative transcriptomic analysis of *S. pneumoniae* strains in response to temperature changes, considerably extending the observations made by Aprianto *et al* (Aprianto *et al.* 2018). During translocation of the pneumococcus from the nasopharynx to the blood to cause invasive pneumococcal disease (IPD), or directly to the lung to cause pneumonia, *S. pneumoniae* must adapt from temperatures of about 30°C on the nasopharyngeal epithelial surface to body temperatures of around 37°C (Weiser, Ferreira, and Paton 2018; Kwon *et al.* 2003). With the aim of understanding how *S. pneumoniae* adapts during this transition, we used RNA-Seq to compare the transcriptomic changes of *S. pneumoniae* serotype 1 and serotype 6B strains during transfer from 30°C to 37°C at multiple timepoints. As discussed earlier, *S. pneumoniae* serotype 6B strains are frequent colonisers of the nasopharynx of healthy individuals but less commonly cause IPD (Brueggemann *et al.* 2004; Weinberger *et al.* 2009). Serotype 1 *S. pneumoniae* on the other hand, are highly invasive strains that have a short duration colonising the host nasopharynx and are frequently associated with epidemics (Brueggemann *et al.* 2004; Weinberger *et al.* 2010; Ruckinger *et al.* 2009; Kwambana-Adams *et al.* 2016).

Serotype 1 grew faster and reached a higher maximum density than serotype 6B at 30°C (**Figure 7.6**). We speculate that, if this is also the case in the host, this faster growth rate of serotype 1 may achieve quicker activation of host defences leading to inflammation, thereby facilitating *S. pneumoniae* transmission (Weight *et al.* 2018) and accelerating translocation to disease-causing niches (Aguilera and Lenz 2020). Hijacking the host immune responses is a common phenomenon in *S. pneumoniae* pathogenesis (Weiser, Ferreira, and Paton 2018; Aguilera and Lenz 2020). Within this context, the slow growth of the serotype 6B strain, if this also occurs in the host, may allow for a longer carriage duration and survival under the radar of the host immune system. In their analyses of *S. pneumoniae* growth under different environmental exposure, in contrast to our results here, Tothpal and colleagues did not observe growth difference in *S. pneumoniae* serotypes grown at different temperatures, (Tothpal *et al.* 2019). However, they used the enriched BHI media, which is optimised to allow for maximal growth of *S. pneumoniae*

(Sanchez-Rosario and Johnson 2021). The ability of the CDM used in this thesis to distinguish growth phenotypes of serotype 1 and serotype 6B supports its use over enriched media for comparative transcriptomic experiments focused on understanding pneumococcal adaptation.

From the controlled experiments, we showed that a change in temperature leads to broad reprogramming of the *S. pneumoniae* transcriptome affecting diverse metabolic pathways. There was a common core set of pneumococcal genes in serotype 1 and serotype 6B strains, about 50% ( $n \sim 260$ ) of all differentially regulated genes, that were regulated during temperature change from 30°C to 37°C. Only six of the nineteen “temperature responsive” genes that were previously identified in a *S. pneumoniae* D39 transcriptomics study at elevated temperatures were found to be upregulated in the present study (Aprianto *et al.* 2018) (**Figure 7.13A**). In our clinical strains, most of these “temperature responsive genes” exhibited both strain dependent and time dependent regulation with four of the genes not differentially regulated by either of our clinical strains (**Figure 7.13A**). For example, the glutamine ABC transport *glnQ2*, which is part of the upregulated temperature responsive genes in D39 (serotype 2) (Aprianto *et al.* 2018), was downregulated by both serotype 1 and serotype 6B clinical strains. These results could reflect differences between serotype 2 (D39) and our strains but could also result from the extensive laboratory adaptation undergone by D39 since its isolation in 1914 (Griffith 1928; Lanie *et al.* 2007), underlining the need for caution in extrapolating experimental findings from laboratory strains. Since we used the same CDM and growth conditions for the temperature shift experiment as Aprianto *et al.*, the observed differences in the regulation of these temperature responsive genes were most likely to reflect genuine strain specific adaptive differences, rather than artefactual differences in experimental conditions. The downregulation of the oxidative stress response genes, including pyruvate oxidase (*spxB*) (**Figure 7.13B**), in elevated temperatures has been previously observed in a microarray gene expression analysis of the TIGR4 clinical strain (Pandya *et al.* 2005). The TIGR4 strain (serotype 4) is an invasive clinical strain which was used to generate the first complete genome sequence of *S. pneumoniae* (Tettelin *et al.* 2001). Intriguingly, some genes upregulated by TIGR4 at elevated temperatures in the microarray study such as *proWX*, *proV*, and *cbpA* were uniquely upregulated by our



serotype 1 strain but not serotype 6B in the present dataset, potentially indicating expression markers for invasiveness or differential pathogenicity (Pandya *et al.* 2005). Similar to their study, bacterial central metabolism and amino acid synthesis were among the most significantly upregulated pathways at elevated temperatures (Pandya *et al.* 2005). The only other 'omics study to investigate pneumococcal response to elevated temperature following an extensive literature search was performed by proteomics (MALDI-TOF) using the D39 (serotype 2) strain following temperature change from 37°C to 42°C (Myoung-Ro 2006). Although only 25 proteins were identified to be differentially regulated at the protein level in this study, likely due to the low resolution of MALDI-TOF methods used, many of the proteins identified in that study such as *spxB* (pyruvate oxidase), *adhE* (alcohol dehydrogenase E), *grpE* (chaperone protein), and *dnaK* (chaperone protein) were among the most significantly regulated genes in our dataset (**Figure 7.13A** and **Figure 7.13B**) (Myoung-Ro 2006). In summary, this approach corroborated differentially regulated genes that had been identified in previous 'omics studies and revealed novel differentially regulated genes that provide novel insights into pneumococcal adaptation to temperature.

To further demonstrate that the differentially regulated genes identified were specific to the change in temperature and not simply a response to stress in general, we assessed the expression of transition metal response genes identified in Chapter 6 in response to iron levels and found that they were mainly downregulated or not differentially regulated (**Figure 7.13C**). The only exception to this in our temperature experiment was *copA* gene, which showed strain specific regulation, being upregulated in serotype 6B but not serotype 1 at both the 10 minutes and 20 minutes timepoints (**Figure 7.13C**). *copA* is part of a three-gene pneumococcal operon that is involved in regulation of copper levels in *S. pneumoniae*, is considered a virulence factor for pneumococcal colonisation, and we have showed in Chapter 6 that expression of *copA* was upregulated by both strains during iron limitation (Shafeeq *et al.* 2011; van Opijnen and Camilli 2012). The *copA* gene was not previously found to be upregulated by RNA-Seq analysis for the D39 strain at elevated temperatures (Aprianto *et al.* 2018). As far as we are aware, the involvement of *copA* in pneumococcal temperature homeostasis, as suggested by our results for serotype 6B strain but not serotype 1, has not been previously reported and requires further

validation. Similar to the results of Pandya *et al* using microarrays (Pandya *et al.* 2005) and Myoung-Ro *et al* using proteomics (Myoung-Ro 2006), we found several kinases and phosphatases to be differentially upregulated by *S. pneumoniae* in response to elevated temperatures. In the present dataset, kinases and phosphatase genes were among the core differentially regulated genes by both serotype 1 and serotype 6B, suggesting a role for phosphorylation in pneumococcal adaptation to temperature change (**Table 7.6A**, **Table 7.6B**, **Table 7.7A**, and **Table 7.7B**) (Shi *et al.* 2014).

Pneumococcal two-component systems pathways were significantly downregulated by both serotype 1 and serotype 6B after 20 minutes in elevated temperature (**Figure 7.11B**). At the single gene level however, the two-component signal transduction system genes *comCDE*, which controls competence state of *S. pneumoniae* through phospho-regulation (Sanchez *et al.* 2015; Martin *et al.* 2010; Guiral *et al.* 2006; Alloing *et al.* 1998; Iannelli, Oggioni, and Pozzi 2005; Steinmoen, Knutsen, and Havarstein 2002), were upregulated by serotype 6B but not serotype 1 at elevated temperatures (**Figure 7.13B**). This result implies increased transformation capacity by serotype 6B as a pathogenic principle during exposure to increased temperatures. The bacterial chaperone and serine protease, high temperature requirement A (*htrA*), was among the most significantly downregulated genes by both serotype 1 and serotype 6B in response to elevated temperatures. *htrA* is a highly conserved protein in both prokaryotes and eukaryotes (including humans and plants) and is being investigated as a potential drug target to prevent bacterial infections (Parui *et al.* 2022; Wessler, Schneider, and Backert 2017). In bacterial pathogens such as *Escherichia coli* (Lipinska *et al.* 1989), *Campylobacter jejuni* (Brondsted *et al.* 2005), and *Streptococcus pyogenes* (Jones *et al.* 2001), where *htrA* is an extracellular protein, knockout mutant studies found *htrA* to be essential for virulence and bacterial thermal stress response (Xue *et al.* 2021). In *S. pneumoniae*, *htrA* contributes to bacterial virulence and knockout mutants are attenuated during colonisation of murine models (Ali *et al.* 2020; Ali *et al.* 2021; Ibrahim *et al.* 2004; de Stoppelaar *et al.* 2013). Given previous reports on the importance of the *htrA* for bacterial adaptation to high temperatures, we expected the expression of the gene to be upregulated following a shift to higher environmental temperatures. Similar to our results, the *htrA* gene as well as its suggested regulators *ciaRH* were also significant downregulated after transfer from 30°C to elevated

temperatures of 37°C in *S. pneumoniae* D39 (serotype) strain (Sebert *et al.* 2002; Aprianto *et al.* 2018). One plausible explanation comes from the findings of Ali *et al.* (2021), where they concluded that pneumococcal serine proteases are significant contributors to pneumococcal colonisation but have a diminished effect in invasive disease (Ali *et al.* 2020). In this context, it is reasonable to expect the downregulation of *htrA* when transferred from a temperature of 30°C, which is the environmental temperature during colonisation, to higher temperature of 37°C which mimics core body temperature. Nonetheless, *htrA* may well be essential to *S. pneumoniae* adaptation to increased environmental temperatures as previously suggested, it is possible that the regulation of the gene in either direction is important, and there may be further regulation at the post-transcriptional level. Therefore, further analysis of *htrA* gene and the mechanisms by which it contributes to pneumococcal virulence need to be elucidated, including the operon structure and potential regulatory mechanisms.

In conclusion, we have established a comparative transcriptomics methodology which has generated a novel pneumococcal transcriptomics dataset and provided new insights into how *S. pneumoniae* adapts its gene expression during exposure to elevated temperatures. The results in this work showed that about 50% of differentially regulated genes in response to elevated temperature may be serotype independent. However, many of the genes that were uniquely upregulated by one strain were still present in the other, supporting a model where differences in strain adaptation, possibly invasiveness and carriage efficiency are due, at least in part to differences in the regulation of core pneumococcal genes. Future work beyond this thesis, will include supplementing the gene expression data described here and in Chapter 6, with dRNA-Seq transcript mapping data, which has been generated as part of this work. This will inform us on the position of TSS and TTS, thus improving our understanding of operon structures and gene regulation for the conditions investigated. The repertoire of small and non-coding RNA in clinical strains of *S. pneumoniae* serotype 1 and serotype 6B are yet to be described. In the TIGR4 and D39 strain for which sRNA and ncRNA data have been generated from RNA-Seq datasets, considerable strain differences were observed with several indicated to be involved in virulence (Sinha *et al.* 2019). Transcript mapping and sRNA results could also be used to

improve the annotation of the complete genome assemblies of serotype 1 and serotype 6B generated in Chapter 4.

## Chapter 8 : Development of techniques for isolation of *S. pneumoniae* RNA from human samples for transcriptomic analysis

## 8.1 Introduction

In Chapters 5, 6, and 7, we have used *in vitro* models to gain insights into biology of *S. pneumoniae* in controlled experimental environments. *In vitro* experiments together with *in vivo* murine studies have been vital in *S. pneumoniae* research helping identify several important bacterial virulence factors and understanding of host-pathogen interactions (Berry, Lock, *et al.* 1989; Berry and Paton 1996; Berry, Yother, *et al.* 1989; Brown, Gilliland, and Holden 2001; Rosenow *et al.* 1997). Findings from these studies have been instrumental in development of effective *S. pneumoniae* interventions such as vaccines and will remain a fundamental part in expanding our knowledge of *S. pneumoniae* biology. Despite the incredible insight they provide about *S. pneumoniae* biology, *in vitro* and murine studies have drawbacks that limit direct extrapolation of findings from experimental settings to the multi-factorial human niche. For example, most *in vitro* experiments are accomplished by growing bacteria in enriched media like Todd-Hewitt broth with yeast (THY) and brain heart Infusion (BHI) broth which are extremely nutritious by designed to maximally support bacterial growth (Sanchez-Rosario and Johnson 2021; Todd and Hewitt 1932). However, such growth environments are not reflective of the conditions the bacteria encounters in the host which is known to be restricted in nutrients the bacteria utilise for growth and survival (Man, de Steenhuijsen Piters, and Bogaert 2017; Weiser, Ferreira, and Paton 2018). Chemically defined media (CDM) such as the nose mimicking condition (NMC) media used in Chapter 6 and Chapter 7, which are more limited in nutrient composition are now more frequently being used to reflect to some extent the host restricted nutrient environment *in vitro* (van de Rijn and Kessler 1980; Sicard 1964; Aprianto *et al.* 2018). To account for the impact of the host immune system on *S. pneumoniae* pathogenesis in experimental laboratory settings, primary and secondary epithelial cell lines and murine models are regularly used to study host-pathogen interactions and understanding the influence of host environments. Cell lines seldom contain the complete repertoire of the human immune cells and can be prone to changing genetic and phenotypic properties over time (Capes-Davis *et al.* 2019; Kaur and Dufour 2012). For example, Mucins which are important host defence component and considered a major source of carbohydrates for *S. pneumoniae* survival during colonisation, are not secreted in the Detroit 562 pharyngeal cell lines when cultured in liquid-liquid interface (Weiser, Ferreira, and Paton 2018). Murine studies are important

for identifying *S. pneumoniae* virulence factors by gene knockout experiments and are a critical part of primary studies in evaluating and advancing potential vaccine candidates (Denoel *et al.* 2011; Adamou *et al.* 2001; Ramos-Sevillano, Ercoli, Felgner, *et al.* 2021; Briles *et al.* 2000; Verhoeven, Xu, and Pichichero 2014). However, some findings in murine models have not been replicated in humans and several pneumococcal vaccine strategies shown to be efficacious in murine models have proven to be ineffective when advanced to human subjects (Odotola *et al.* 2017; Hammitt *et al.* 2019). In addition to not being the natural host of *S. pneumoniae*, murine models exhibit immunological and anatomic differences in host niches such as the upper respiratory tract (URT) compared to humans (Chiavolini, Pozzi, and Ricci 2008; Seok *et al.* 2013; Yue *et al.* 2014). Furthermore, murine models are inherently immune to certain *S. pneumoniae* strains that cause disease in humans and the strains that can cause disease in murine models require non-physiological infection doses (Chiavolini, Pozzi, and Ricci 2008; Wade *et al.* 2014; Briles *et al.* 1992).

The limitations of experimental models emphasise the importance of developing the ability to study the biology of *S. pneumoniae* in its natural niche, the human upper respiratory tract (URT) (Bogaert, de Groot, and Hermans 2004; Weiser, Ferreira, and Paton 2018). The Experimental Human Pneumococcal Carriage (EHPC) model, set up at the Liverpool School of Tropical Medicine (LSTM), United Kingdom, provides an important platform for studying *S. pneumoniae* biology in its natural niche (Wright *et al.* 2012; Ferreira *et al.* 2013; Gritzfeld *et al.* 2013). In the EHPC model, healthy adults are recruited and intranasally challenged with live *S. pneumoniae* strains to study carriage dynamics in participants and to understand factors that affect colonisation. The reports from the EHPC studies have given important insights into immunological manifestations during *S. pneumoniae* colonisation (Jochems *et al.* 2019; Mitsi *et al.* 2020; Wright *et al.* 2012). However, the reports are lacking in approaches that use advanced molecular biology and bioinformatics methods to investigate *S. pneumoniae* adaptations in the human nasopharynx. Transcriptomic analysis of *S. pneumoniae* from its natural niche in the URT has been a challenge predominantly due to the lack of methods for obtaining high-quality bacterial RNA from this environment. Recent advances in next generation sequencing have led to more sensitive methods such as dual-RNA sequencing that requires lower concentration of starting RNA material compared to earlier methods (Mardis 2017;

Poulsen and Vinther 2018; Croucher and Thomson 2010; van Vliet 2010; Sharma and Vogel 2014; Colgan, Cameron, and Kroger 2017; Westermann, Gorski, and Vogel 2012). Dual RNA-seq allow for simultaneous investigation of transcriptome of both pathogen and host cells and has been successfully applied for bacterial pathogens *Salmonella* (Westermann *et al.* 2016; Aulicino *et al.* 2022; Westermann, Gorski, and Vogel 2012), *Escherichia coli* (Wang *et al.* 2020), *Mycobacterium tuberculosis* (Pisu *et al.* 2020; Lopez-Agudelo *et al.* 2022), and *S. pneumoniae* (Aprianto *et al.* 2016; Minhas *et al.* 2020). However, all these studies were performed using cell lines in *in vitro* settings. Even in these *in vitro* settings, bacterial RNA quantities can be small which translates to high sequencing cost due to the requirement for high sequencing depth to generate enough reads for differential gene expression analysis. Bacterial mRNA is less than 2% of total RNA in samples from infected host cells (Betin *et al.* 2019; Westermann, Gorski, and Vogel 2012). Recently described genome enrichment strategies have been shown to increase the proportion of bacterial mRNA in these samples by a factor of greater 10x thus requiring shallower sequencing depth to acquire enough reads for differential expression analysis (Betin *et al.* 2019; Aulicino *et al.* 2022).

In this chapter, a proof-of-principle study was conducted to demonstrate isolation of high-quality RNA from human URT suitable for *S. pneumoniae* transcriptomic analysis. In Chapter 2, an optimised RNA isolation protocol was described which was shown in Chapter 6 and Chapter 7 to be efficient for obtaining high quality *S. pneumoniae* RNA in *in vitro* conditions. Leveraging samples collected in EHPC model in Liverpool, we hereby evaluate the efficiency of our RNA isolation protocol using human samples. The EHPC samples used for the current work were collected from participants challenged with *S. pneumoniae* BHN418 (serotype 6B) strain, which is the same strain for which we have generated transcriptomic data in response to iron and temperature changes in Chapter 6 and Chapter 7 respectively. Successful generation of transcriptomic data from these human samples will therefore enable comparison with *in vitro* datasets and help inform how well the *in vitro* findings reflect transcriptomic adaptations of *S. pneumoniae* in human URT. More broadly, our methods may be an important step in enabling the wider scientific community to study transcriptomic adaptation of *S. pneumoniae* in its natural niche in the human URT.



## 8.2 Methods

### 8.2.1 Experimental human pneumococcal carriage (EHPC) samples

Six oropharyngeal samples (OPS) and nasal fluid samples (NFS) were collected from healthy adults who participated in the controlled Experimental Human Pneumococcal Carriage (EHPC) study at Liverpool School of Tropical Medicine (LSTM), United Kingdom (Wright *et al.* 2012; Gritzfeld *et al.* 2013). The EHPC study recruitment protocol, inclusion and exclusion criteria, ethics approval, and samples collection and processing has been described in detail previously (Ferreira *et al.* 2013; Gritzfeld *et al.* 2013; Wright *et al.* 2012). Healthy adults were intranasally challenged with a single dose of serotype 6B (strain BHN418) which was boosted by another intranasal challenge of the same strain 14 days later. *S. pneumoniae* carriage was monitored by nasal wash on days 2, 7, 16, and 21 post challenge. *S. pneumoniae* colonisation in challenged individuals was confirmed at LSTM by positive culture using classical microbiology methods (Gritzfeld *et al.* 2013). Microbiology assessments of carriage of *S. pneumoniae* was performed using nasal fluid samples. Details of the six EHPC samples collected for the current study including *S. pneumoniae* colonisation status is presented in **Table 8.1**. The OPS samples collected from the six participants including the sampling swabs were placed in 0.5mL RNA Later and stored at -80°C. Transport of the samples from LSTM was completed overnight in dry ice and storage maintained at -80°C until processing.

### 8.2.2 RNA isolation of nasopharyngeal samples (NPS) retrieved from healthy volunteers

The EPHC samples were limited and considered too valuable for the initial optimisation of the RNA isolation protocol. Therefore, additional nasopharyngeal samples (NPS) were collected from the nose of two healthy adult male volunteers (volunteer 1 and volunteer 2) from University College London (UCL), United Kingdom for purposes of optimising RNA isolation protocol. Sample collection from the volunteers was performed by trained personnel and placed in 1mL RNA preservation media and stored at -80°C. Consent was obtained from volunteers' and samples were collected and processed in line with UCL ethics and data protection policies. NPS samples were collected because this was what the laboratory personnel were trained to collect and because a recently collected NPS of

one of the volunteers (volunteer 1) was determined to be positive for *S. pneumoniae* in a separate study.

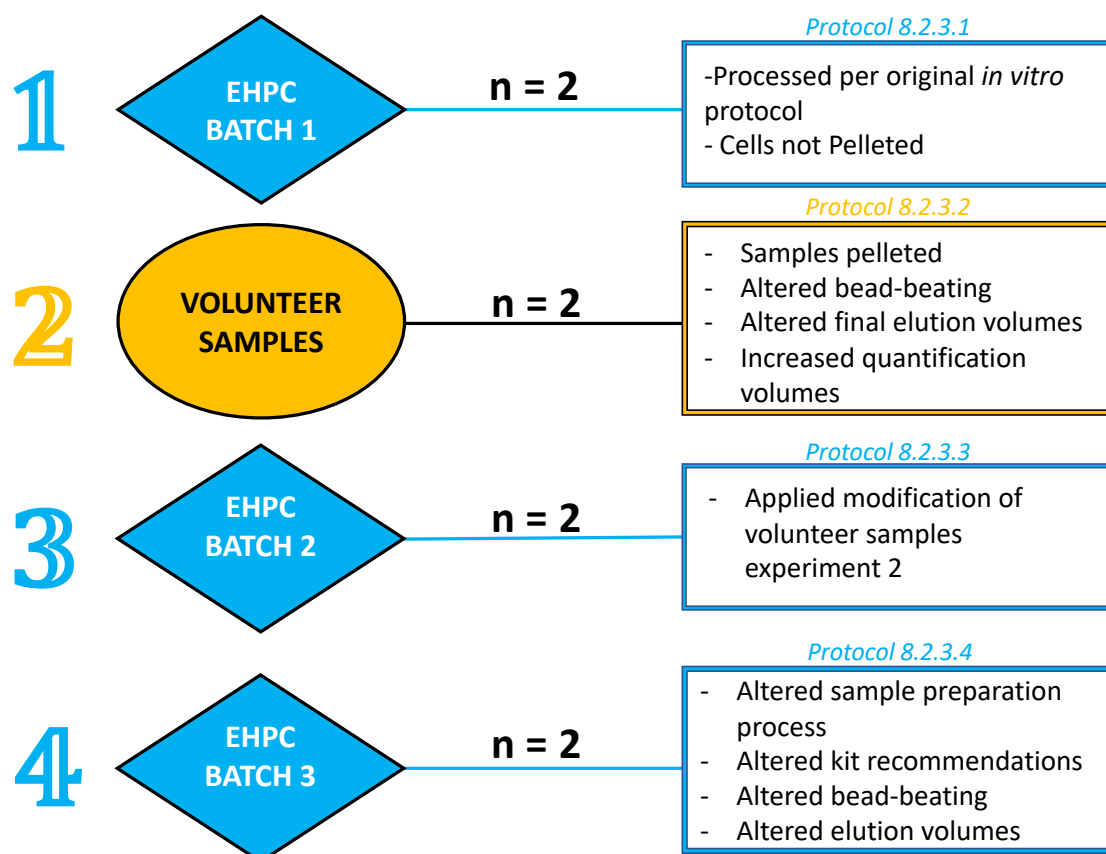
**Table 8.1:** Six oropharyngeal samples (OPS) collected from participants in the Experimental Human Pneumococcal Carriage (EHPC) study. Pneumococcal carriage was determined from nasal wash samples by conventional microbiology methods.

Sample ID	<i>S. pneumoniae</i> carriage	OPS collection day
3528/029/028	Positive	D16
3531/029/031	Negative	D16
3533/029/033	Positive	D7
3534/029/034	Negative	D7
3538/029/038	Positive	D21
3540/029/040	Positive	D7

### 8.2.3 RNA extraction of EHPC samples

The original RNA isolation protocol used has been described in detail in Chapter 2. The original protocol has been optimised for *in vitro* experiments (Chapters 6 and Chapter 7). To evaluate the protocol for *in vivo* samples and to allow for further optimisation, the six EHPC samples were divided into 3 different batches (2 samples each) and RNA was isolated separately with minor changes to the original protocol. The changes to the original protocol applied to each batch were intended to sequentially optimise and improve RNA yield from each batch of samples. Generally, the MirVana miRNA extraction kit (Invitrogen) was used, and lysis of cells was performed by addition of 1mL of lysis buffer supplemented with 50U mutanolysin (Sigma) and 30mg/mL lysozyme (Sigma). Samples in lysis buffer were placed in 2mL lysing tubes containing glass beads (Precelleys) and placed in tissue homogenizer (Precelleys). Samples were bead beaten at 6200rpm for 45 seconds 3x within 20 seconds intervals. The homogenate was placed at 70°C for 10 minutes and then centrifuged at 4°C for 5 minutes at 12,000g. The supernatant was placed in QIAshredder (Qiagen) and centrifuge at maximum speed for 2 minutes. The resulting lysate was used to isolate RNA following manufacturer's protocol with final elution

volume of 100µL. Extracted RNA was DNA digested using the TURBO DNA-free kit (Invitrogen) following manufacturer's protocol. RNeasy mini-elute clean-up kit (Qiagen) was finally used to clean the RNA samples with final elution volume of 22µL. The modifications applied to original RNA extraction protocol for each batch is summarised in **Figure 8.1** and sections 8.2.3.1 – 8.2.3.4 below.



**Figure 8.1:** Schematic summarising sequence of experiments and modifications to the RNA isolation protocol applied to each experiment. Six oropharyngeal samples (OPS) from experimental human pneumococcal carriage (EHPC) study were divided into three batches and isolated separately to sequentially improve RNA yield. Nasopharyngeal samples (NPS) from adult volunteers were also included for optimisation purposes. Experiments are sequentially indicated in numeric. The section numbers in the current document detailing protocol modifications for each experiment is indicated. Blue colours indicate experiments from EHPC samples while gold colour indicate volunteer samples experiment.

#### **8.2.3.1 RNA isolation protocol modifications: EHPC samples batch 1**

EHPC Samples 3528/029/028 and 3531/029/031 make up the first batch. The original RNA isolation protocol was used for this batch with the only modifications being related to samples processing. Here, lysis buffer was directly added to OPS samples including preservative media (RNA Later) and then processed as per the original RNA isolation protocol. For *in vitro* experiments, sample preparation involves pelleting bacterial cells before lysis. The rationale for this modification was that very few cells were expected to be recovered from the EHPC samples and it was presumed that centrifugation would not generate bacterial pellets. It was also presumed that bacterial RNA may be lost by discarding supernatant as part of pelleting process.

#### **8.2.3.2 RNA isolation protocol modifications: non EHPC volunteer samples**

The second RNA isolation modifications were applied to the two NPS samples collected from volunteer 1 and volunteer 2. In comparison to the isolation of the first batch of EHPC samples, the following modifications were made for this experiment:

- Squeezed sampling swabs against the storage tube walls to extract as many cells/tissues as possible.
- Pelleted the samples by centrifugation at 12,000g for 10 minutes and discarded supernatant.
- Added lysis buffer to pellets, similar to *in vitro* samples processing.
- Increased bead-beating from x3 for 45 seconds at 6400rpm to x5 for 45 seconds at 6400rpm to enhance mechanical cells lysis.
- Reduced final elution volume in clean-up stage with RNeasy mini-elute kit (Qiagen) from 22µL to 10µL to increase RNA concentration.

#### **8.2.3.3 RNA isolation protocol modifications: EHPC samples batch 2**

Samples 3538/029/038 and 3540/029/040 make up the second batch of EHPC samples. For this experiment, the modifications applied to the RNA isolation protocol for the volunteer NPS samples (section 8.2.3.2) were replicated.

#### **8.2.3.4 RNA isolation protocol modifications: EHPC samples batch 3**

The final isolation of EHPC samples were for samples 3533/029/033 and 3534/029/034.

The following modifications were adopted for this experiment.

- More emphasis was placed on squeezing as much of the cells/tissues from the swabs as possible. Here, we scraped the sampling swabs with mild force between the lid and tip of sterile 1.5mL tubes.
- Bead beating was reverted to x3 for 45 seconds at 6400rpm. The additional bead beating (x5 for 45 seconds at 6400rpm) was unlikely to have added value based on results of the previous experiment.
- Eliminated one of the additional washing steps in MirVana miRNA extraction kit (Invitrogen). The rationale was to preserve more RNA in the columns during elution process.
- Final elution with MirVana miRNA extraction kit (Invitrogen) was performed twice using 100µL of elute each time (200µL total). Previously, elution was done once as per manufacturer protocol with 100µL. The rationale was to increase yield by retrieving more RNA from column.
- The elution volume of RNeasy mini-elute kit (Qiagen) was changed to 15µL which was closer to manufacturer's recommendation (14µL). Originally, 22µL was used to elute RNA but this was modified to 10µL in volunteer experiments and for EHPC batch 2 experiments above. Elution was done once with RNeasy mini-elute kit (Qiagen).
- 2µL of isolated RNA was used for Qubit assay to enable quantification/detection by the Qubit fluorometer.

#### **8.2.4 RNA quantification**

Quality and quantity of total RNA was assessed using Nanodrop spectrometer (Thermo), Qubit assay kit (Thermo), and Bioanalyzer system (Agilent) as described in Chapter 2.

### 8.2.5 Confirmation of *S. pneumoniae* by qPCR

To confirm the presence of *S. pneumoniae* in the isolated RNA samples, quantitative and block PCR reactions were performed based on previously described methods targeting 3 different conserved *S. pneumoniae* genes (Carvalho Mda *et al.* 2007). The targeted pneumococcal genes were (i) autolysin (*lytA*), (ii) pneumolysin (*ply*) and (iii) serotype 6B capsule. The *lytA* and *ply* PCR assays were used for species-specific detection of *S. pneumoniae* while serotype 6B assay was used to confirmed whether positive specimens are of the challenged EHPC serotype 6B. The primer sequences for each of the three assays is listed in **Table 8.2** and based on recommended methods from the [Centre for Disease Control and Prevention \(CDC\)](#).

**Table 8.2:** Primers and probes used for *S. pneumoniae* PCR detection targeting conserved pneumococcal genes.

PCR Target	Primer/Probe	Sequence (5' – 3')
6B capsule	Forward	AATTTGTATTTTATTCATGCCTATATCTGG
	Reverse	TTAGCGGAGATAATTTAAAATGATGACTA
<i>lytA</i>	Forward	ACGCAATCTAGCAGATGAAGCA
	Reverse	TCGTGCGTTTTAATTCCAGCT
	Probe	TGCCGAAAACGC" T" TGATACAGGGAG
<i>ply</i>	Forward	GCTTATGGGCGCCAAGTCTA
	Reverse	CAAAGCTTCAAAAGCAGCCTC
	Probe	CTCAAGTTGGAAACCACGAGTAAGAGTGATGAA

#### 8.2.5.1 cDNA generation from RNA samples

Isolated RNA samples were first reverse transcribed (+RT) to cDNA using LunaScript® RT SuperMix Kit (New England BioLabs). The reverse transcriptase protocol was as follows: Reaction mixes were made to a total volume of 20µL in nuclease free PCR tubes. Each reaction contained 4µL of 5X LunaScript master mix, 1µL RNA, and 15µL RNase-free water. The LunaScript kit (New England BioLabs) includes a non-RT (-RT) master mix which was used as negative control of reverse transcriptase step. A non-template control (NTC) was

also included as an additional control. The reactions were run in a thermocycler with the following cycling conditions: 25°C for 2 minutes, 55°C for 10 minutes and 95°C for 1 minute. The resulting cDNA was used for qPCR reactions or kept in -80°C for long term storage.

#### **8.2.5.2 *S. pneumoniae* PCR detection using serotype 6B specific primers**

The Luna Universal qPCR Master mix kit (New England Biosciences) was used to set-up qPCR with serotype 6B specific primers (**Table 8.2**). For each sample technical triplicates for the +RT and technical duplicates for the -RT samples were used. A primer master mix with final concentration of 10µM of each of the forward and reverse primers was prepared separately. Each well of qPCR plate contained 20µL total reaction master made of 10µL of 2X Luna qPCR master mix, 1µL of primer master mix (final concentration of 250nM/primer), 8µL of sterile water, and 1µL of appropriate cDNA template (cDNA +RT and cDNA -RT). PCR plates were covered with optical adhesive film to prevent evaporation and briefly centrifuged at 1000g for 30 seconds. The plate was placed in Eppendorf quantitative qPCR equipment using the following cycling conditions: 95°C for 5min, followed by 40 cycles of 95°C for 15secs and 60°C for 45 secs. A block PCR was also set up using the CDC serotype 6B primers (**Table 8.2**) to allow for visualisation of amplicons by gel electrophoresis. Each reaction of the block PCR contained 12.5µL of PCR mastermix, 1.25µL of primer mastermix, 6.25µL of sterile water, and 5µL of cDNA. The cycling conditions for the block PCR were a cycle of 94°C for 4 minutes, followed by 30 cycles at 94°C for 45 seconds + 54°C for 45 seconds + 72°C for 2 minutes, and a cycle at 72°C for 2 minutes.

#### **8.2.5.3 *S. pneumoniae* detection targeting autolysin (*lytA*) and pneumolysin (*ply*) genes**

Specific detection real-time PCR detection requires additional probes that specifically target sequences in addition to amplification primers. This method (also known as fluorescence resonance energy transfer (FRET)) increases specificity of amplified targets and is more sensitive than non-specific detection systems. Primer and probe working solutions of 2.5µM concentration were prepared to be directly added to PCR assay master mix. Assays were carried out in volumes using the PerfeCTa® qPCR ToughMix (Quanta bio)

according to manufacturer's instructions. Each reaction mix in PCR plate wells was 23µL in total with reagent volumes shown in **Table 8.3** for *lytA* and *ply* qPCRs. Each sample was run as triplicates. 3µL water was added to non-template control (NTC) wells. RNA isolated from *S. pneumoniae* strain BHN418 (serotype 6B) grown to mid-exponential phases (MEP) in THY *in vitro* was used as positive controls. The DNA sample used for whole genome sequencing of BHN418UCL (Chapter 4) was also included as a positive control in these assays. Plates were sealed with plastic adhesive seal and centrifuged at 1000g for 30 seconds. Plates were placed in Eppendorf qPCR thermocycler with the following cycling conditions: a cycle at 50°C for 2 minutes, a cycle at 95°C for 10 minutes, and 45 cycles at 95°C for 15 seconds + 60°C for 1 minute.

**Table 8.3:** qPCR mastermixes for *lytA* and *ply* assays

Reagent	1x (µL)	12x (µL)
<i>lytA</i> qPCR assay		
PerfeCTa master mix	12.5	150
<i>lytA</i> Forward primer	0.5	6
<i>lytA</i> Reverse primer	0.5	6
<i>lytA</i> Probe	0.5	6
Water	6	72
<i>ply</i> qPCR assay		
PerfeCTa master mix	12.5	150
<i>ply</i> Forward primer	1	12
<i>ply</i> Reverse primer	1	12
<i>ply</i> Probe	1	1
Water	4.5	54

### 8.2.6 Data analysis

Amplification plots and plate sample values generated by qPCR machine was used to determine amplification cycles numbers for each sample. Amplification plots were normalised by subtracting baseline signal of fluorescent dye and determining the fluorescent threshold. The cycle number at which sigmoidal amplification curves for each



sample crosses the fluorescent threshold was referred to as the cycle threshold (Ct) and used to determine sample positivity. Samples with Ct  $\leq 36$  were considered positive and samples with Ct  $> 40$  were considered negative. Samples with Ct of 37-39 were considered equivocal.

### 8.2.7 Gel electrophoresis

The amplicon of PCR amplification products for serotype 6B specific block PCR assay was visualised using agarose gel electrophoresis. 10  $\mu$ L of PCR reaction product was mixed with 2  $\mu$ L of 6X loading dye and loaded on the gel. 4  $\mu$ L of 1Kb DNA size marker was loaded with samples to estimate amplicon sizes. The agarose gel concentration used was 1.5% in Tris/acetate/EDTA (TAE) buffer and ran at 100 volts for 40 minutes. Gels were visualised under UV light.

## 8.3 Results

### 8.3.1 Bacterial quantities retrieved from EHPC participants

Using classical microbiology methods, our collaborators at LSTM showed that pneumococcal carriage was variable between the six EHPC participants following *S. pneumoniae* challenge and administration of booster challenge (**Table 8.4**). For participants that became colonised, bacterial load (carriage density) was evaluated by calculating the number of colony forming units (CFU) at days 2, 7, 16, and 21. Four of the six EHPC participants had positive *S. pneumoniae* culture results on days when OPS samples were collected and therefore determined to be carriage positive individuals (**Table 8.4**). The quantity of bacteria in challenge inoculums as well as the amount retrieved from nasal washes of the six participants on each of the studied days is summarised in **Table 8.4**.

**Table 8.4:** Bacteria quantities (CFU) inoculated and retrieved for six EHPC participants. Participants were intranasally challenged with *S. pneumoniae* strain BHN418 (serotype 6B) on day 1 and day 14. *S. pneumoniae* carriage was evaluated on day 2, 7, 16, and 21.

Sample	Inoculation CFU/100µL	D2 CFU/mL	D7 CFU/mL	D14 booster (CFU/100µL)	D16 CFU/mL	D21 CFU/mL	OPS
3528/029/028	87,834	8.819	76.599	83,667	3.912		D16
3531/029/031	74,667		203.752	90,333			D16
3533/029/033	89,000	0.506	2.783	90,000	9.719	30.641	D7
3534/029/034	83,834	9.298		89,334			D7
3538/029/038	87,834		2.115	98,333	12.333	1126.781	D21
3540/029/040	92,834	37.119	745.588	98,000	1465.283	31178.225	D7

### 8.3.2 RNA extraction results: EHPC samples batch 1

The RNA quantification results for EHPC samples 3528/029/028 and 3531/029/031 isolated with protocol modifications stated in section 8.2.3.1 are presented in **Table 8.5**. RNA yield from these EHPC samples was considerably lower compared to *in vitro* experiments (see Chapter 6 and Chapter 7) where RNA yield was between ~100ng/µL and >1400ng/µL. Due to the low quantity of RNA from these samples, dilution factor used for qubit quantification was below the instrument limit of detection. To improve RNA yield, further optimisation of RNA isolation protocol was conducted using the non-EHPC samples collected from healthy adult male volunteers.

**Table 8.5:** RNA quantification results of batch 1 EHPC samples.

Sample ID	Nanodrop			Bioanalyzer (ng/µL)	Qubit (ng/µL)
	ng/µL	260/280	260/230		
3528/029/028	5.5	2.35	0.7	3	Too Low
3531/029/031	7	2.26	0.98	21	Too Low

Too low = RNA Quantity was below limit of detection

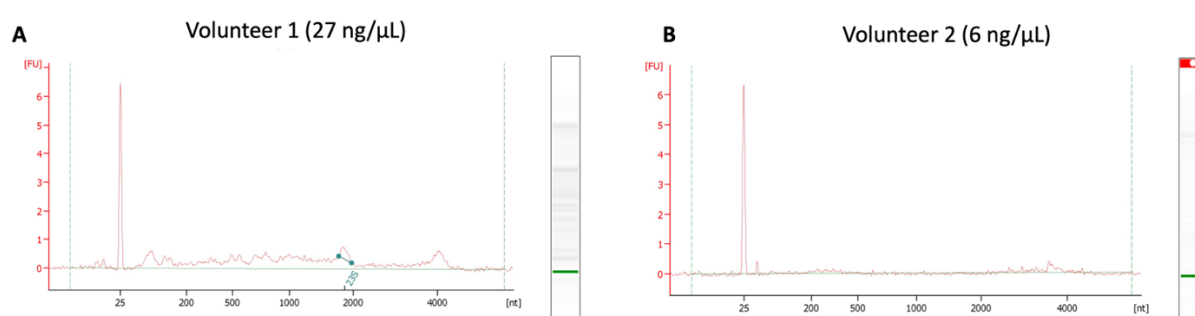
### 8.3.3 RNA extraction results: volunteer samples

The RNA quantification results following isolation of RNA from nasopharyngeal samples (NPS) of volunteer 1 and volunteer 2 for the purposes of optimisation are presented in **Table 8.6**. The modifications (Section 8.2.3.2) to the isolation protocol applied here resulted in increased RNA quantities from 5-7ng/ $\mu$ L to 27ng/ $\mu$ L for volunteer 1 compared to results for EHPC batch 1 RNA results (**Table 8.6**). RNA bands for 16S and 23S for volunteer 1 were clearly visible from the bioanalyzer gel image (**Figure 8.2**) indicating intact microbial RNA. The poor RNA yield in volunteer 2 was attributed to poor sample collection. In addition, the carriage status of volunteer 2 was unknown. The protocol modifications used in this experiment were determined to have improved RNA quantities and applied to the next batch of EHPC samples.

**Table 8.6:** RNA quantification results of NPS obtained from volunteers (optimisation).

Sample ID	Nanodrop			Bioanalyzer	
	ng/ $\mu$ L	260/280	260/230	ng/ $\mu$ L	RIN
Volunteer 1	18.2	2.15	0.77	27	1.4
Volunteer 2	1.7	2.86	0.07	6	N/A

N/A= Software cannot calculate RIN



**Figure 8.2:** Electropherogram and gel image of total RNA isolated from NPS samples of two volunteers for the purpose of optimising RNA isolation protocol. Volunteer 1 (A) showed high-quality RNA with clear 16S and 23S bacteria ribosomal peaks indicating intact RNA. Volunteer 2 (B) had lower quality RNA, and this was attributed to poor sample collection.

### 8.3.4 RNA extraction results: EHPC samples batch 2

The modifications that improved RNA yields in volunteer samples batch were applied to EHPC samples 3538/029/038 and 3540/029/040. The RNA quantification results of this experiment are presented in **Table 8.7**. There was modest improvement in RNA yield compared to first batch of EHPC samples, but the RNA concentrations were lower than those obtained for volunteer 1. Increasing the dilution factor for qubit instrument enabled generating RNA quantification results for EHPC batch 2 samples. At this juncture, it was reasoned that the lower RNA yield from the EHPC samples compared to volunteer 1 results may have been due to differences in sample type (NPS vs OPS).

**Table 8.7:** RNA quantification results of batch 2 EHPC samples.

Sample ID	Nanodrop			Bioanalyzer		Qubit
	ng/ $\mu$ L	260/280	260/230	ng/ $\mu$ L	RIN	ng/ $\mu$ L
3538/029/038	5.8	2.61	0.02	14.0	1.0	6.4
3540/029/040	6.9	2.22	0.23	10.0	N/A	8.7

### 8.3.5 RNA extraction results: EHPC samples batch 3

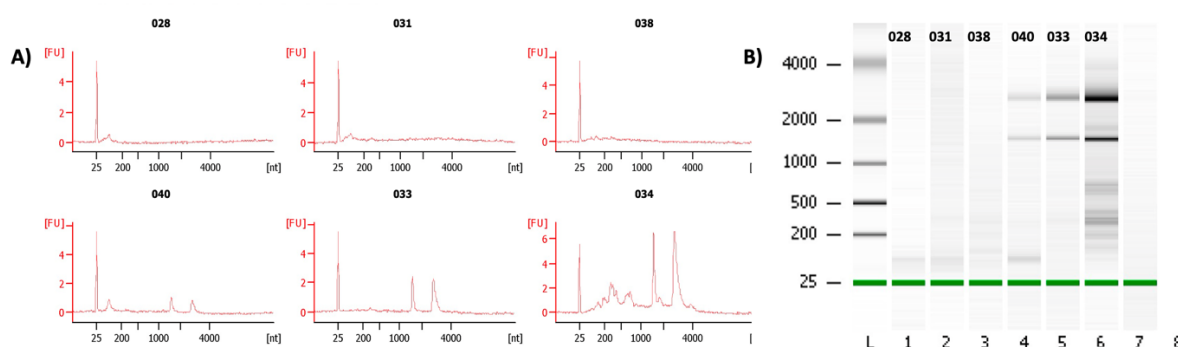
For the isolation of the final two EHPC samples, several modifications related to sample processing and manufacturer recommendations were adopted (see section 8.2.3.4). These changes resulted in best RNA quantification results for the EHPC samples 3533/029/033 and 3534/029/034 (**Table 8.8**). High yield RNA with RNA integrity number (RIN) as high as those attained for *in vitro* experiments (>7.5) (Chapter 6 and Chapter 7) were obtained for these EHPC samples. Although one of the EHPC samples in this batch (3534/029/034) was not colonised by *S. pneumoniae* as per conventional microbiology results (**Table 8.1**), the sample contained the highest yield of total RNA of all EHPC samples with distinct 16S and 23S bands indicative of presence of microbial pathogen (**Figure 8.3**). The presence of *S. pneumoniae* in all the isolated RNA of EHPC samples was therefore further evaluated using molecular biology detection methods.

**Table 8.8:** RNA quantification results of batch 3 EHPC samples

Sample ID	Bioanalyzer		Qubit (ng/ $\mu$ L)	Nanodrop (ng/ $\mu$ L)
	ng/ $\mu$ L	RIN		
3533/029/033	16	8.9	10.8	9.9
3534/029/034	80	7.5	56.3	57.4

### 8.3.6 Modifications to RNA isolation protocol improved RNA quality of EHPC samples

The Bioanalyzer allows for RNA integrity checks and produces electropherogram and gel images of RNA samples. The Bioanalyzer electropherogram and gel image results showed that the last three isolated EHPC samples had the best RNA quantity and quality suggesting sequential improvements in RNA yield because of optimisation steps (**Figure 8.3**). The RIN scores obtained in isolation of the last batch of EHPC samples was in concordance with those obtained for *in vitro* experiments.

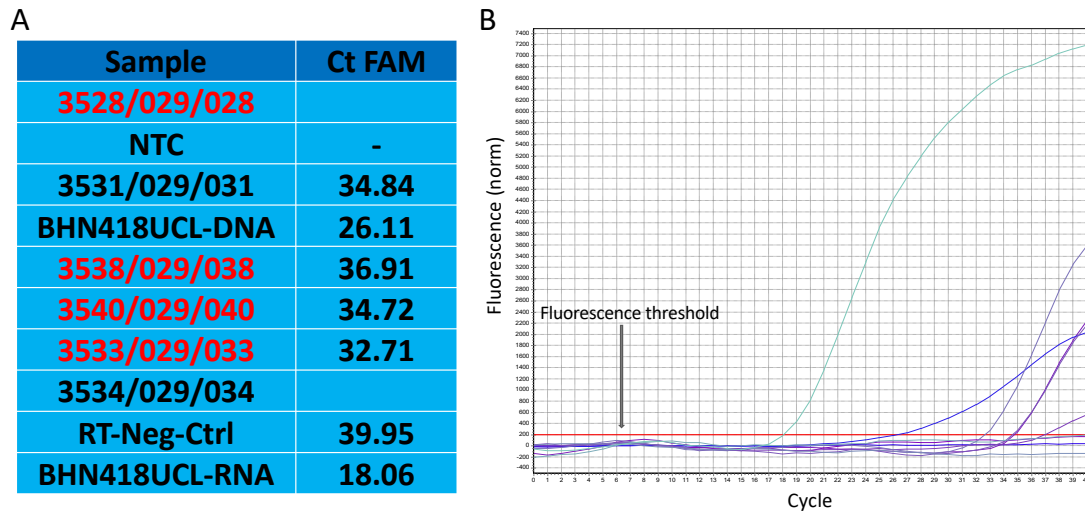


**Figure 8.3:** Bioanalyzer chip-based capillary electrophoresis of RNA from six EHPC samples. RNA quantification was done with Agilent RNA 6000 Nano Reagents using the Prokaryote Total RNA Series II module. **A.** The electropherogram results of peaks showing 16S and 23S peaks in three samples (040, 033, and 034). **B.** The corresponding gel images of the six EHPC samples. Samples in lanes 4-6 (040, 033, and 034) of the gel show distinct bands representing 16S and 23S RNA. The sample IDs are annotated on images.

**8.3.7 *S. pneumoniae* detection in EHPC RNA by quantitative PCR targeting *lytA***

*lytA* qPCR identified four of the six EHPC RNA samples to be positive for *S. pneumoniae*. The presence or absences of *S. pneumoniae* by *lytA* qPCR in four samples was consistent with results of conventional microbiology (**Table 8.1**). The discrepant results were for 3528/029/028 which was reported as positive by conventional microbiology methods but found to be negative by *lytA* qPCR (Ct >40) and 3531/029/031 which was reported as negative by classical microbiology but found to be positive by *lytA* qPCR (Ct =34.84) (**Table 8.1** and **Figure 8.4**). Mid-exponential phase (MEP) *in vitro* samples of BHN418UCL used as positive controls were amplified as expected with BHN418UCL RNA (cDNA) having Ct of ~18 and DNA of the strain having a Ct of ~26 (**Figure 8.4**). The EHPC sample with the most abundant prokaryotic RNA quantity (3534/029/034) was negative for *S. pneumoniae* by qPCR corroborating classical microbiology results and indicating specificity of the qPCR assay (**Table 8.8** and **Figure 8.4**). The average Ct value for *S. pneumoniae* positive samples by *lytA* qPCR was ~34.

## *S. pneumoniae* detection using *lytA* primers



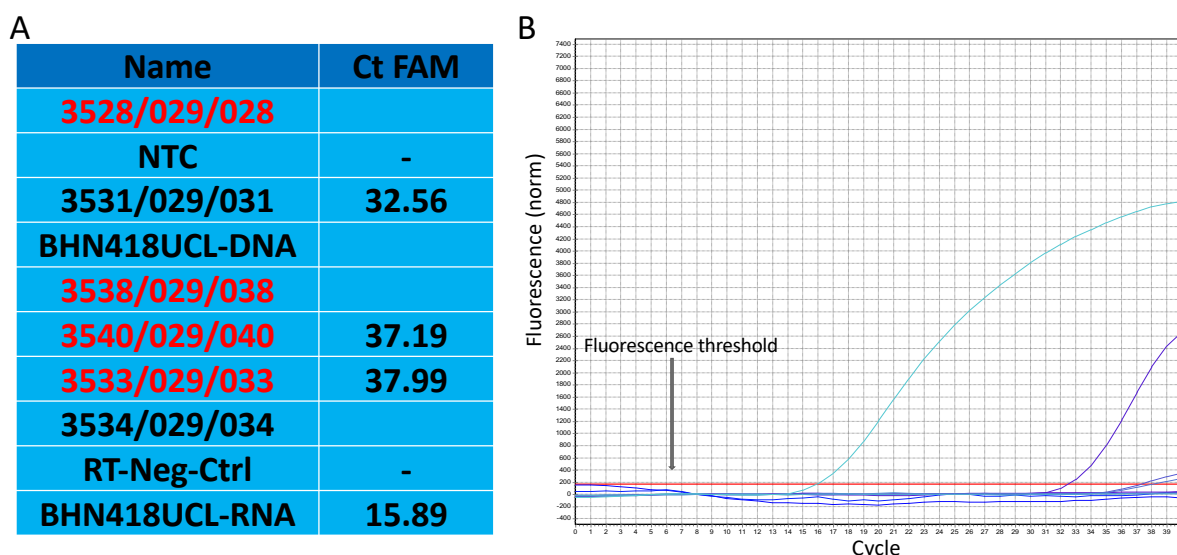
**Figure 8.4:** *S. pneumoniae* detection in EHPC samples by TaqMan qPCR targeting *lytA* (autolysin) gene. **A.** Sample description and cycle threshold (Ct) values obtained for each reaction. IDs in red colour have been determined by conventional microbiology methods to be positive for *S. pneumoniae*. The DNA and cDNA samples of BHN418UCL (serotype 6B) were used as positive controls for qPCR assay. **B.** qPCR amplification plot of all samples highlighting fluorescence threshold used as cut-off.

**8.3.8 *S. pneumoniae* detection in EHPC RNA samples by quantitative PCR targeting *ply***

*ply* qPCR also identified 3531/029/031 as *S. pneumoniae* positive with Ct of ~32 (**Figure 8.5**). The 3531/029/031 sample was the discrepant results between *lytA* qPCR and classical microbiology results which may be explained by higher sensitivity of PCR methods. In addition, differences in bacterial amounts in NFS used for classical microbiology and OPS used for qPCR may cause the discrepant results. Although the transcribed BHN418UCL RNA (cDNA) positive control (Ct ~15) and negative controls (NTC and -RT) were amplified as expected, the Ct values of other *S. pneumoniae* positive EHPC samples were relatively high (Ct ~37) (**Figure 8.5**). Ideally these results would have been repeated, potentially with further optimisation of assay. However, due to the limited amount of RNA samples left for subsequent work, the *ply* qPCR assay was not repeated. Furthermore, *lytA* qPCR for which optimal data was obtained (**Figure 8.4**), is the

recommended PCR detection assay for *S. pneumoniae* because it best separates it from other closely related streptococci species (Messmer *et al.* 2004; Neeleman *et al.* 2004; Carvalho Mda *et al.* 2007).

## *S. pneumoniae* detection using *ply* primers



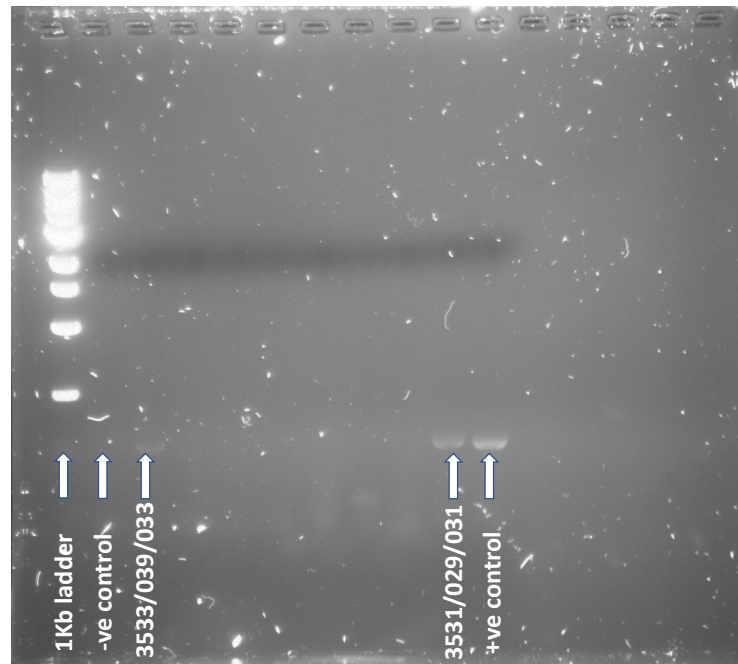
**Figure 8.5:** *S. pneumoniae* detection in EHPC samples by TaqMan qPCR targeting *ply* (pneumolysin) gene. **A.** Sample description and cycle threshold (Ct) values obtained for each reaction. IDs in red colour have been determined by conventional microbiology methods to be positive for *S. pneumoniae*. The DNA and cDNA samples of BHN418UCL (serotype 6B) were used as positive controls for qPCR assay. **B.** qPCR amplification plot of all samples highlighting fluorescence threshold used as cut-off.

### 8.3.9 Confirmation of *S. pneumoniae* serotype 6B by PCR

Using *S. pneumoniae* serotype 6B specific primers, amplicons were obtained for 3531/029/031 and 3533/029/033 confirming that these participants were colonised by the serotype of *S. pneumoniae* that was used for challenge in EHPC model (**Figure 8-6**). The data confirms that 3531/029/031 is indeed a carrier of BHN418, corroborating *lytA* and *ply* qPCR results and contradicting conventional microbiology data. PCR amplicons were not visible by gel electrophoresis for the two other positive samples (3538/029/038



and 3540/029/040) likely due to low cycle numbers used for block PCR (30 cycles). As per qPCR results, the average Ct for detection of *S. pneumoniae* in these EHPC samples is ~34. Further optimisation of this assay was not performed due to low RNA volumes which was preserved for sequencing.



**Figure 8.6:** *S. pneumoniae* serotype specific block PCR detection for serotype 6B. RNA samples from two EHPC participants (3531/029/031 and 3533/039/033) were confirmed to be carriers of the serotype that was used to challenge EHPC participants.

## 8.4 Discussion

The failure of some *S. pneumoniae* vaccine candidates that have been selected from *in vitro* and murine studies to be immunogenic in humans may be linked to differences in *S. pneumoniae* adaptation and gene expression compared to human host (Odutola *et al.* 2017; Hammitt *et al.* 2019; Hakansson 2014). In this chapter, we aimed to isolate high-quality *S. pneumoniae* RNA from samples retrieved from the human URT for bacterial transcriptomics analysis. Following multiple rounds of optimisation, we have developed a protocol for isolating *S. pneumoniae* RNA from URT of human carriers. As the samples were collected from participants that were challenged with a known strain of *S. pneumoniae* (serotype 6B, strain BHN418), we were able to confirm the presence of that specific serotype using molecular detection methods. These results present an opportunity to use RNA-Seq to attempt a global transcriptome profiling of *S. pneumoniae* from its natural niche, the human URT.

One of the EHPC participants that was reported as a negative carrier by conventional microbiology methods was found to be colonised by the serotype of BHN418 strain here, suggesting higher sensitivity of molecular detection methods (Selva *et al.* 2013; Almeida *et al.* 2020). However, we could not confirm the presence of *S. pneumoniae* in one sample that was reported to be positive by conventional microbiology methods. The reason for the discordance in the sample which could not be confirmed by molecular methods may be related to differences in bacterial densities between nasal fluid samples (NFS) and oropharyngeal samples (OPS). Conventional microbiology culture methods were used to assess carriage of *S. pneumoniae* in NFS at LSTM. Here, we used PCR to identify *S. pneumoniae* from OPS samples collected from the same individuals. Discrepancies in *S. pneumoniae* presence between the two sampling methods (OPS vs NPS) has been previously noted by the LSTM group (personal communication) and this may explain the discordant results in the two samples.

RNA-Seq provides quantification of gene expression to the single nucleotide level and has become the method of choice for transcriptomics studies due to its high resolution (Westermann, Gorski, and Vogel 2012; Poulsen and Vinther 2018). However, based on the published literature, the application of this technology in studying transcriptome of

bacterial pathogens in human host has been unsuccessful thus far. The hinderances include the inability to isolate enough pathogen RNA to enable sequencing at enough depth for transcriptomics analysis (Westermann, Gorski, and Vogel 2012). Previous attempts to investigate *S. pneumoniae* transcriptome in EHPC participants at LSTM were unsuccessful due to poor RNA quantification results (Daniela Ferreira personal communication). The NanoString nCounter system provides an alternative for directly measuring individual mRNA transcripts from mixed cultures including human samples (Geiss *et al.* 2008; Barczak *et al.* 2012; Moffitt *et al.* 2021; Basset *et al.* 2022). However, the resolution of NanoString is several orders of magnitude lower than RNA-Seq and is only capable of targeting a few hundred predefined mRNAs, limiting chances of discovering novel gene expression patterns (Geiss *et al.* 2008; Tsang *et al.* 2017; Eastel *et al.* 2019). In addition, data normalisation methods, which are critical for gene expression analysis have not been standardised for NanoString data (Wang *et al.* 2016; Molania *et al.* 2019; Bhattacharya *et al.* 2021). The lowest amount of total RNA in *S. pneumoniae* positive samples in the present study was 10ng/ $\mu$ L (Bioanalyzer results of sample 3540/029/040). Using a novel strategy in their *in vitro* study of *Salmonella* and host transcriptomic responses, Aulicino and colleagues (Aulicino *et al.* 2022) successfully conducted transcriptomic analysis from samples with total RNA concentrations as low as 4.17ng/ $\mu$ L, supporting the feasibility of transcriptomics analysis from our human samples.

In the next phase of this work, beyond this thesis, we propose to use the SureSelect (Agilent) target enrichment library preparation method which has been successfully used to enriched pathogenic sequences in host specimens (Aulicino *et al.* 2022; Clark *et al.* 2018). Enrichment of bacterial mRNA prior to RNA-Seq reduces sequencing cost and removes unwanted bacterial RNA species while maintaining core biological signals (Betin *et al.* 2019; Aulicino *et al.* 2022). Our probe design strategy will use the genome sequence of the BHN418 strain which we have re-sequenced in Chapter 4 (BHN418UCL) to design target probes that exclude genomic sequences encoding tRNAs and rRNAs. We will also BLAST probe sequences against the human genome to identify and remove highly similar sequences to the BHN418 strain. At the library preparation stage, we will deplete host rRNA from the samples to prevent potential interference with probe specificity. Probes will then be used to enriched *S. pneumoniae* mRNA in the four EHPC positive samples

identified in the present study. We plan to include at least four replicates for each sample to improve statistical power to identify differential gene expression (DGE). Each library will be sequenced at an average depth of 20 million reads with reads length of 75-150bp, which has been demonstrated to be adequate for DGE analysis in enriched libraries (Betin *et al.* 2019; Aulicino *et al.* 2022). To investigate DGE, we will compare the expression of each EHPC samples to expression of mid-exponential phase growth of BHN418 in THY *in vitro*.

The present study may have limitations that were not apparent from previous studies that employed similar strategies (Betin *et al.* 2019; Aulicino *et al.* 2022). First, the human URT is resident to several other bacterial and viral pathogens which may have highly similar orthologs to *S. pneumoniae* (Man, de Steenhuijsen Piters, and Bogaert 2017; Donati *et al.* 2010). Therefore, despite the specificity of our probe design for *S. pneumoniae* strain BHN418 (serotype 6B), we may not be able to rule out the enrichment and expression of transcripts from other (unknown) commensals of the oral cavity. *Streptococci* species such as *S. mitis* and *S. oralis* are recognized to be involved in horizontal gene transfer with *S. pneumoniae* in the URT and share several genes (Lessa *et al.* 2018; Kalizang'oma *et al.* 2021; Donati *et al.* 2010). Secondly, our proposed experimental strategy may not be able to account for the contribution of host differences (e.g., the immune system) that may influence the bacterial transcriptome. *In vitro* studies are not confounded by this issue since standardised single host cell types are usually infected with known strains at specific multiplicity of infection (MOI). Despite the potential limitation in data interpretation, it would be an important first step to be able to demonstrate generation of enough pathogen reads for DGE analysis of *S. pneumoniae* obtained from the human URT as a proof-of-principle.

## Chapter 9 : General discussion

### 9.1 Summary and general discussion

Advances in next generation sequencing methods, such as high throughput genome-wide quantification of transcripts, have helped unravel important biological insights into how bacterial pathogens adapt gene expression in a conditional manner (Nicolas *et al.* 2012; Kroger *et al.* 2013). However, most of our knowledge on bacterial gene regulation has been derived from a few model microorganisms such as *Escherichia coli* and *Bacillus subtilis*. Given the requirement of *S. pneumoniae* to adjust to different environmental conditions during colonisation, transmission and disease, RNA-mediated gene regulation likely plays a key role in allowing the bacteria to quickly adapt in new environments. It is therefore surprising that studies on RNA based gene regulation of the important bacterial pathogen *Streptococcus pneumoniae*, using high-resolution RNA sequencing methods have been scarce (Wilton *et al.* 2015; Aprianto *et al.* 2018; Warriar *et al.* 2018).

*S. pneumoniae* was identified as a causative agent of pneumonia, meningitis, and sepsis more than 150 years ago (Pasteur 1881; Sternberg 1881; Fraenkel 1886; Klemperer and Klemperer 1891a; Griffith 1928). However, fundamental biological gaps remain in our understanding of how a commensal that normally colonises the human URT, promotes transmission and adapts to other bodily environments, potentially causing fatal disease. It is also not fully understood why, among the 100 different serotypes identified so far, certain pneumococcal serotypes are more likely to cause disease than others. Serotype 1, for example, is recognised as one of the most invasive pneumococcal strains (Chaguza *et al.* 2016; Ritchie, Mitchell, and Evans 2012; Williams *et al.* 2012). Indeed, in the African meningitis belt, *S. pneumoniae* serotype 1, although less prominent than *Neisseria meningitidis*, is a not infrequent cause of meningitis epidemics (Greenwood 1999; Kwambana-Adams *et al.* 2016; Leimkugel *et al.* 2005; Parent du Chatelet *et al.* 2005). In contrast, other serotypes such as serotype 6B are frequent colonisers of the URT but less commonly cause disease (Weinberger *et al.* 2009; Brueggemann *et al.* 2004). Bacterial genetics are thought to account for about 63% of phenotypic variance such as carriage duration and disease development due to *S. pneumoniae* (Lees *et al.* 2017). Host traits on the other hand supposedly account for less than 5% of the phenotype highlighting the importance of *S. pneumoniae* bacterial factors (Lees *et al.* 2017). However, variations in the genome alone does not explain the differences between strains in their ability to cause

disease, with about 75% of genes being in the core genome, including a range of virulence factors (Donati *et al.* 2010). A recent GWAS analysis of the highly invasive serotype 1 strains recovered from carriage and invasive disease host niches globally did not identify discernible differences in genetic content between strains of the same serotype recovered from the different niches (Chaguza *et al.* 2022). Given that for other bacterial pathogens such as *Salmonella* (Ono *et al.* 2005; Hammarlof *et al.* 2018), *Escherichia coli* (Santos-Zavaleta *et al.* 2019) and *Bacillus subtilis* (Nicolas *et al.* 2012), where regulation of gene expression is a key driver of pathogenesis, it may be that pneumococcal disease progression is not just the presence or absence of specific bacterial genes. Aprianto *et al.* have reported the most comprehensive transcriptomic profiling of *S. pneumoniae* to date using RNA-Seq to investigate pneumococcal gene regulation under different environmental exposures (Aprianto *et al.* 2018). This study highlights the highly dynamic nature of pneumococcal gene regulation and provides a platform for novel hypotheses testing. However, like many studies in pneumococcal research, the authors used the D39 (serotype 2) laboratory-adapted strain (Slager, Aprianto, and Veening 2018). The D39 strain, which is highly virulent in murine models, was first isolated in 1914, and has been continually passaged ever since (Griffith 1928; Slager, Aprianto, and Veening 2018). The advantage of the continued use of D39 has mainly revolved around making results obtained from different laboratories comparable. However, the public health translational potential of these results may be limited. Laboratory-adapted strains like D39 and R6 (an unencapsulated variant of D39) have acquired a range of genetic traits which have been shown to affect key metabolic pathways including virulence genes (Lanie *et al.* 2007). The high rate of new *S. pneumoniae* interventions shown to be efficacious in murine models that subsequently fail in human studies, may in part be due to the use of laboratory strains in the primary studies (Odutola *et al.* 2017). The main goal of this thesis was therefore to utilise RNA-Seq-based transcriptomic analysis of clinically relevant *S. pneumoniae* strains to understand how the bacteria adapts to various biologically relevant conditions. Hence, the following were achieved during this process:

- i) Identification of suitable growth media and experimental methods to conduct controlled transcriptomic studies of clinical *S. pneumoniae* strains (**Chapter 3**).

- ii) Generation and characterisation of high-quality genome sequences of *S. pneumoniae* clinical strains of serotype 1 and serotype 6B for transcriptomic analysis (**Chapter 4**).
- iii) Development of a bioinformatics pipeline for the analysis of *S. pneumoniae* RNA-Seq data (**Chapter 5**).
- iv) Characterisation of the transcriptomic adaptation of clinical *S. pneumoniae* serotype 6B mutants lacking biosynthesis genes *fhs* and *proABC* during growth under stress conditions (**Chapter 5**).
- v) Conduct of the first comparative transcriptomics analysis of clinical *S. pneumoniae* strains of serotype 1 and serotype 6B in response to iron (**Chapter 6**) and temperature (**Chapter 7**) changes over time.
- vi) Exploration of the feasibility of advancing *S. pneumoniae* transcriptomics studies to human carriage (**Chapter 8**) by demonstrating the successful isolation of high-quality *S. pneumoniae* RNA directly from human oropharyngeal samples.

In **Chapter 3**, we identified the most suitable CDM for subsequent transcriptomics experiments following a trial of several different growth media described in the literature. Results in this chapter showed that the selected CDM was able to distinguish the growth phenotype of different pneumococcal strains based on differences in carbohydrate source, temperature, and iron concentrations. Carbohydrate source is a requirement for survival of *S. pneumoniae* (Weiser, Ferreira, and Paton 2018) and different carbohydrate sources potentially impact pneumococcal capsule production differently (Troxler *et al.* 2019). Growth analysis in Chapter 3, showed that growth in galactose and glucosamine, which are the two most abundant carbohydrates in the human nasopharyngeal surface, the natural niche of pneumococcus, best supported the growth of pneumococcal strains and grew to higher maximal growth density compared to other carbohydrates such as sucrose and fructose (Paixao, Oliveira, *et al.* 2015). However, all pneumococcal strains tested retain the ability to grow in the various carbohydrate sources used. *S. pneumoniae* encodes at least 10 extracellular glycosidases that allow it to modify many glycans (King 2010) and dedicates about 30% of its genome to biosynthesis and transport of various carbohydrates (Bidossi *et al.* 2012; Tettelin *et al.* 2001). We identified both carbohydrate



source and temperature related differences in growth between pneumococcal strains with different carriage or invasiveness potential. During growth in galactose, glucosamine, and sucrose, the high-carriage frequency, less-invasive serotype 6B strain grew faster and to a higher maximum density than the lower-carriage frequency, higher-invasive TIGR4 strain (serotype 4). Although *S. pneumoniae* strains were found to grow to a lower maximal growth density in fructose compared to the other carbohydrates tested, which is consistent with previous studies (Troxler *et al.* 2019; Weinberger *et al.* 2009), growth in fructose showed a reversed pattern with the more invasive serotype 4 (TIGR4) strain growing faster and to a higher growth density than the serotype 6B (BHN418) strain. Capsule thickness of *S. pneumoniae* has been shown to significantly reduced during growth in fructose compared to growth in glucose, with more heavily encapsulated strains such as serotypes 6B, which has a capsule size three times thicker than serotype 4, being more affected (Weinberger *et al.* 2009). Differences in the intracellular concentrations of phosphorylated metabolites such as UDP-glucose and UDP- galactose, which are important precursor molecules for capsule production, have been implicated in phenotypic growth differences of *S. pneumoniae* in fructose (Troxler *et al.* 2019). Thus, the absence of precursor metabolites for capsule production during growth in fructose likely inhibits *S. pneumoniae* ability to produce capsule, depriving the bacteria of an important carbon source for metabolism and leading to reduced growth as observed in our growth experiments. We speculate that the metabolism of the serotype 6B strain, which has a thicker capsule than serotype 4 strain (TIGR4) and therefore likely requires higher amount of precursor molecules for capsule production, is more significantly affected by this lack of energy source and leading to a slower growth rate than serotype 4 (TIGR4) during growth in fructose. The transcriptomic basis of the differential effect of different carbohydrate sources on growth of *S. pneumoniae* was not investigated in this thesis and remains to be investigated.

We had hypothesized that predominantly carriage strains like serotype 6B are better adapted to the conditions of the nasopharyngeal environment and would have a different growth kinetic compared to strains with higher invasive potential like serotype 1 and serotype 4 at 30°C. Therefore, we compared the growth of strains at temperatures similar to colonisation of the nasopharyngeal epithelial surfaces (30°C), core body temperature

(37°C), and body temperature during high fever (39°C). Although growth of serotypes 1, 4, and 6B were similar at 37°C and 39°C, at 30°C the more invasive strains of serotype 1 and serotype 4 grew faster and to a higher maximal growth than serotype 6B. If these experiments reflect what happens *in vivo*, further experiments investigating the underlying pneumococcal adaptation to these conditions may explain the difference in carriage and invasiveness potential of strains.

We have also showed as part of work undertaken in Chapter 3 that chelation of growth media with 2,2'-bipyridyl (BIP), which reduces the concentration of extracellular transition metal ions, impaired pneumococcal growth in a dose-dependent manner with increasing concentration of BIP increasingly leading to growth defects in serotype 1, serotype 4, and serotype 6B strains. The impaired growth phenotypes associated with chelation of transition metal ions was rescued by addition of iron to growth medium. The different niches *S. pneumoniae* occupies in the host are known to vary in concentration of transition ion metals with the colonisation niche, the nasopharynx, having lower levels of free ion compared to the bloodstream (McDevitt *et al.* 2011; Honsa, Johnson, and Rosch 2013). Iron is an important co-factor for many bacterial processes and our *in vitro* results indicate that extracellular iron levels directly influence pneumococcal growth potentially through acceleration/deceleration of metabolic processes.

In addition to identifying the appropriate growth medium and characterising the growth phenotypes of different *S. pneumoniae* strains under specific environmental conditions, critical optimisation work was undertaken in Chapter 3 for subsequent transcriptomics experimental. These included optimisations of RNA isolation protocol to generate high-quality RNA suitable for RNA-Seq, selection of suitable RNA stabilisation reagents to arrest bacterial transcription at completion of experiments, identification of mid exponential growth phase of different strains under experimental condition, and showing that bacterial optical density, which was used as a proxy for pneumococcal growth, correlated with number of live bacteria as measured by CFU/mL. These optimisation experiments ensured design of robust experiments to investigate transcriptomics response of *S. pneumoniae* to changes in iron and temperature in later chapters.

RNA sequencing requires a high-quality reference genome sequence for accurate mapping and quantification of transcripts. However, by the start of this project, less than 1% of the approximately 10,000 publicly available pneumococcal genome assemblies were complete. In **Chapter 4**, we leveraged the combined strengths of long-read (PacBio) and short-read (Illumina) sequencing to generate contiguous *de novo* genome assemblies of *S. pneumoniae* clinical strains of serotype 6B (BHN418UCL) and serotype 1 (BVJ1JL) (Wick *et al.* 2017; Betts *et al.* 2021). To ensure accuracy of genome assemblies, both strains were sequenced to very high genome depth which were in excess of 500x. Comparative genomic analysis of the generated assemblies indicated that approximately 75% (1,674 genes) of the overall pangenome were present in both of the clinical strains of serotype 1 and serotype 6B used. In addition to supporting our subsequent transcriptomic analysis, generating high-quality assemblies provide a valuable resource in the wider scientific community (Betts *et al.* 2021).

The goal in **Chapter 5** was to develop a transcriptomic analysis pipeline for differential gene expression (DGE) and identification of biological pathways from *S. pneumoniae* transcriptomic data. To do this, we utilised the transcriptomic dataset generated by Dr Elisa Ramos-Sevillano at UCL of *S. pneumoniae* serotype 6B single gene mutants grown under stress conditions. Single gene knockout mutagenesis studies are important in understanding bacterial genotype-phenotype relationships and have been instrumental in the identification of key *S. pneumoniae* virulence factors. However, our analysis in Chapter 5 indicated that single gene mutations may have broader effects on the pneumococcal transcriptome than previously thought, with potentially wide-ranging regulatory consequences. Transcriptomic profiling of serotype 6B (BHN418) mutants lacking biosynthesis genes *fhs* and *proABC*, which have both been shown to be attenuated in murine models of pneumococcal disease (Ramos-Sevillano, Ercoli, Felgner, *et al.* 2021), revealed considerable differential gene regulation during serum stress. Although the results in this chapter highlighted the classical principle for bacterial compensatory mechanism to gene loss, whereby the mutants adapted to the loss of the knocked-out genes by upregulating the expression of alternate genes and pathways, the upregulation of several off-target genes and pathways unrelated to the deleted biosynthesis genes were also observed. For example, in both  $\Delta fhs$  and  $\Delta proABC$  *S. pneumoniae*, fatty-acid

synthesis (FAS) genes were the most significantly upregulated genes even in culture in human serum where we expected the expression of fatty acid genes to be repressed due to high-lipid content in human serum. Whole cell mutants are being considered as next-generation pneumococcal vaccine strategies, due to potential for broader protection than serotypes specific PCV vaccines (Masomian *et al.* 2020). Preclinical evaluation of  $\Delta fhs$  and  $\Delta proABC$  *S. pneumoniae* have been shown to be attenuated for virulence but are able to colonise the murine nasopharynx as efficiently as wild type strains (Ramos-Sevillano, Ercoli, Felgner, *et al.* 2021). The mutants are being evaluated as potential “harmless” strains to use to deliberately inoculate humans to elicit robust adaptive immunity. However, in addition to the unexplained upregulation of FAS genes during growth in human serum, both mutant strains upregulated the expression of several pneumococcal virulence genes such as *ply*, *psaA*, and *nanA* relative to the wild type strains. The transcriptomics analysis in this work suggests a need for further evaluation of the mutant strains to ensure unintended virulent consequences when used in human participants. This work therefore reveals the importance of pneumococcal metabolism in pneumococcal adaptation and virulence. For the *S. pneumoniae* strain lacking the *fhs* gene, which is involved in the biosynthesis of purine and several amino acids, the strain significant upregulated secondary biosynthesis pathway genes including purine, pyruvate, and quorum sensing genes. The *proABC* genes are involved in pneumococcal proline synthesis. The *proABC* knockout mutant mainly showed upregulation of carbohydrate metabolism and transports genes including galactose and phosphotransferase system pathways. The result in Chapter 5 suggests that the deletion of *fhs* and *proABC* genes in *S. pneumoniae* may cause strain attenuation by changing bacterial metabolism beyond the gene deletions targeted.

In **Chapter 6** we conducted comparative transcriptomics investigation of *S. pneumoniae* during exposure to different iron concentrations while **Chapter 7** focused on pneumococcal transcriptomic adaptation during transfer from temperatures of 30°C to 37°C. In both chapters, we compared the transcriptome of *S. pneumoniae* serotype 1 and serotype 6B strains. Physiological iron concentrations and temperature are two of the most distinct environmental properties between the human URT, which is the

colonisation niche, and other disease-associated niches including the lungs, meninges, and blood. The temperature in the URT is 30°C (Lindemann *et al.* 2002; Kwon *et al.* 2003; Pandya *et al.* 2005; Aprianto *et al.* 2018). As bacteria translocate to other body parts to potentially cause disease, they must adapt to temperatures of at least 37°C (Lindemann *et al.* 2002; Kwon *et al.* 2003; Pandya *et al.* 2005; Aprianto *et al.* 2018). Iron and transition metal ions which are co-factors for many bacterial biological processes, are available in low concentration in URT in contrast to blood which has high iron levels. Our growth results showed that both environmental temperature and iron concentration have significant impact on pneumococcal growth with strain specific differences observed between serotype 1 and serotype 6B. Iron concentration influenced *S. pneumoniae* growth by increasing bacteria growth in iron replete condition and causing growth attenuation in low iron conditions. Transcriptomic analysis of the effect of iron-repletion and iron depletion in Chapter 6 identified several pneumococcal genes that included well known pneumococcal iron response genes including the *piu* and *pia* genes (Brown *et al.* 2001; Brown and Holden 2002). We have also identified several novel putative and hypothetical genes that appear to be involved in iron regulation and may require further validation. Intriguingly, we showed that the differences in transcriptomic adaptation of the highly invasive *S. pneumoniae* serotype 1 strain and less invasive serotype 6B strain is in the expression of core genes. An example of genes that showed strain-specific regulation were the bacteriocin-like peptides (*blp*) genes which were upregulated in serotype 1 in response to iron but not in serotype 6B. Bacteriocins are thought to be expressed by pneumococcus to outcompete other pathogens and we postulate that this may be one of the strategies employed by serotype 1 during its short colonisation duration of the human URT. Overall, the data in chapter 6 validated the involvement of several previously reported genes and gene pathways in *S. pneumoniae* iron homeostasis but challenged other conclusions which may potentially be due to the use of more clinically relevant pneumococcal strains in this thesis. Similarly, some of the genes that were previously shown to be differentially regulated in the D39 (Serotype 2) strain response to elevated temperatures such as *glnQ2*, *clpP*, *clpC*, *cstR*, and *thiX* (Aprianto *et al.* 2018) were shown in **Chapter 7** to not be regulated in the transcriptomic responses of serotype 1 and serotype 6B to increased temperature. About 50% of all genes regulated by serotype 1 and serotype 6B during temperatures change from 30°C to 37°C were

shared. Of the genes that showed strain-specific regulation, the majority were common in both strains, similar to the pneumococcal iron transcriptomic results in Chapter 6. One of the interesting findings in Chapter 7 is the upregulation of competence genes *comCDE*, which controls competence state in *S. pneumoniae* (Pozzi *et al.* 1996; Sanchez *et al.* 2015; Martin *et al.* 2010; Guiral *et al.* 2006), during adaptation to elevated temperature by serotype 6B but not serotype 1. This result suggests that increased transformation capacity may be a feature of *S. pneumoniae* serotype 6B during translocation from the nasopharynx to other body parts but not for serotype 1. The results in chapter 7 also emphasise the importance of chaperone proteins (Samtani, Unni, and Khurana 2022; Zhang *et al.* 2009; Filannino *et al.* 2018) as core bacterial response genes to elevated temperatures which were upregulated by both strains. Kinases and phosphatase genes were found to be part of the core upregulated genes by both serotype 1 and serotype 6B, suggesting the potential role of phosphorylation as a core pneumococcal response mechanism to elevated temperatures (Shi *et al.* 2014; Samtani, Unni, and Khurana 2022).

The work in **Chapter 8** aimed to advance our ability to study *S. pneumoniae* transcriptome directly from human samples, thus addressing a major research gap. The investigation of pneumococcal transcriptome in its natural niche in the host has been mainly hampered by the inability to obtain sufficient high-quality bacterial RNA from host samples. Using the RNA isolation methods developed in this thesis, we established the ability to isolate high-quality pneumococcal RNA from humans colonised by *S. pneumoniae*. Since the samples were retrieved from participants in the experimental human pneumococcal carriage (EHPC) model, who were colonised by a specific strain (serotype 6B, strain BHN418), molecular methods using targeted primers confirmed the presence of the strain used in isolated RNA samples. Subsequent work will utilise novel transcriptomics library preparation methods that enrich bacterial transcripts to allow for gene expression analysis (Betin *et al.* 2019; Aulicino *et al.* 2022). This work has the potential to pave the way for directly studying bacterial adaptation in host samples.

## 9.2 Future Work

A major component of this thesis was the development and refinement of methods for the generation and analysis of *S. pneumoniae* comparative transcriptomic datasets. However, there are multiple potential avenues for future analysis of these data. These are summarised below:

### (i) Improve the genome annotations of *S. pneumoniae* serotype 1 and serotype 6B assemblies

The reference quality genome assemblies of *S. pneumoniae* serotype 1 and serotype 6B generated in Chapter 4 were annotated using automated methods. Manual inspection and correction of hypothetical gene names and gene boundaries will be important. Manual annotation will involve blasting hypothetical protein names on NCBI database to identify evidence for new gene names and correcting these on the annotation where applicable. The PacBio sequencing method provides bacterial epigenetic information which could be incorporated into the updated genome assemblies. Bacterial epigenetic markers are emerging as an important mechanism in bacterial gene regulation. For examples, *S. pneumoniae* has been suggested to regulate colony opacity through epigenetic modifications (Li *et al.* 2019).

### (ii) Identification of transcript boundaries

One of the ways to improve our understanding of *S. pneumoniae* gene regulation is by mapping of transcript start sites (TSS) and transcript termination sites (TTS). For the iron (Chapter 6) and temperature (Chapter 7) comparative transcriptomic datasets, differential RNA-Seq (dRNA-Seq) datasets were generated for all the conditions and timepoints investigated (Sharma and Vogel 2014; Ryan *et al.* 2020). Analysis of the dRNA-Seq dataset will complement the gene expression results presented through identification of novel transcripts, identification of primary and processed transcripts, refining TSS and operon structure, identification of gene promoters and other potential regulators (e.g., small RNAs), identification of strain specific differences in pneumococcal regulation owing to changes in promoter or coding region sequences, etc (Ryan *et al.* 2020; Thomason *et al.* 2015). The information from these analyses will be used to improve the annotation of our reference genome sequences in (i).

**(iii) The small RNA-repertoire of *S. pneumoniae* serotype 1 and serotype 6B clinical strains**

Small regulatory RNAs are involved in the posttranscriptional regulation in bacterial response to environmental changes including virulence (Hammarlof *et al.* 2018). Currently, the *S. pneumoniae* sRNA have been described for the laboratory strain D39 and the TIGR4 clinical strains (Sinha *et al.* 2019). Comparison of TIGR4 and D39 strains sRNA have indicated strain specific difference with up to 15% of sRNA uniquely found in one of the strains. Characterisation of sRNA repertoire of the serotype 1 and serotype 6B strains will expand our knowledge of pneumococcal small RNA regulation and their potential contribution to virulence in these clinical strains (Sastry *et al.* 2019).

**(iv) Application of machine learning techniques**

The use of machine learning tools with biological data is increasing due to non-biased ability to identify biological patterns of relevance from high-dimensional datasets (Bar *et al.* 2021; Poudel *et al.* 2020; Yoo *et al.* 2022; Chou *et al.* 2015; Sastry *et al.* 2021; Niu *et al.* 2019; Naskulwar and Pena-Castillo 2022). Considering the amount of multi-level data being generated in this project, application of unsupervised machine learning tools may, for example further delineate pneumococcal transcriptional regulatory networks and reveal gene co-regulation patterns. In their analysis of pneumococcal RNA-Seq dataset for D39 strain, Aprianto *et al.* employed correlation matrix analysis which helped identify a novel gene that was part of the pneumococcal competence regulon (Aprianto *et al.* 2018).

**(v) Development of a genomic and transcriptomics visualisation browsers**

The establishment of a centralised browser resource of the genomic and transcriptomics data generated in this project would be useful to the wider scientific community. Such a resource will encourage hypothesis generation and may accelerate the pivot to using these clinically relevant *S. pneumoniae* strains more often in laboratory experiments. Transcriptomics databases have become useful resources for hypothesis generation in *Salmonella* (Perez-Sepulveda and Hinton 2018; Kroger *et al.* 2013), *S. pneumoniae* D39 strain (serotype 2) (Aprianto *et al.* 2018), *Escherichia coli* (Chang, Li, *et al.* 2011), *Bacillus subtilis* (Geissler *et al.* 2021) and *Staphylococcus aureus* (Gopal, Nagarajan, and Elasri



2015). In Chapter 5, a prototype browser for the visualisation of the transcriptomic data of the iron experiment was developed.

**(vi) RNA sequencing and transcriptomic analysis of *S. pneumoniae* directly retrieved from human samples**

In chapter 8, we discussed the motivations for developing methods used to successfully isolate *S. pneumoniae* RNA from the nasopharyngeal samples of individuals colonised by *S. pneumoniae* serotype 6B (BHN418). We also described an RNA-Seq library preparation strategy based on enrichment of pathogen transcripts to enable generating enough bacterial reads for gene expression analysis (Betin *et al.* 2019; Aulicino *et al.* 2022). Sequencing and analysis of these samples is ongoing.

### 9.3 Limitations

Transcript abundance is used in RNA-Seq as a proxy for protein abundances but transcript levels by themselves are imperfect predictors of protein abundances with the best current estimates showing a correlation of less than 70-80% (Liu, Beyer, and Aebersold 2016). RNA-Seq elucidates a snapshot of transcript levels at a specific timepoint. To moderate this limitation, we investigated transcript abundance levels at different timepoints for the comparative transcriptomics experiments. Although more timepoints could not be included due to logistic and cost considerations, the present study is the first to investigate the transcriptome of clinical *S. pneumoniae* strains at multiple timepoints. Bacterial pathogens such as *Salmonella* (Kroger *et al.* 2013) and *S. pneumoniae* D39 strain (serotype 2) (Aprianto *et al.* 2018) have been previously shown to quickly change their gene expression upon exposure to new environments with several genes being differentially regulated as early as 5 minutes. In future studies, targeted investigation of the transcriptome at several timepoints and growth stages such as at lag and stationary phase of growth may reveal further insights in bacterial adaptation to new environments. In our analyses, we used *S. pneumoniae* serotype 1 as the high invasive model clinical strain while serotype 6B was used as the less invasive clinical strain. A broader panel of clinical strains categorised as high invasive and less invasive should be included in future experiments. Finally, other physiological conditions that are different in the human

nasopharynx (colonisation niche) and other disease niches like the human lungs, meninges, and blood such as carbohydrate source, pH, and oxidative stress levels (van Opijnen and Camilli 2012; Man, de Steenhuijsen Piters, and Bogaert 2017) could form the basis for future studies.

#### 9.4 Outlook

Bacterial transcriptomic studies have the potential to broaden our understanding of bacterial pathogenesis and can potentially help identify more effective clinical interventions. The results from comparative transcriptomic analysis of serotype 1 and serotype 6B in this thesis indicated that the differences in the regulation of core pneumococcal genes may be a key driver of differences in strain pathogenesis. If these results are validated, interventions targeted to increase or decrease the expression of specific core genes based on contribution to colonisation, transmission, or disease development may be viable. A recent study in plant biology demonstrates the potential utility of RNA-Seq in identifying effective interventions. Wei *et al* (2022) used RNA-Seq to identify genes that contributed to increased yields in rice. Subsequent analysis showed that changing the expression of the most promising candidate, a transcriptional factor, increased rice yield 41-68% (Wei *et al.* 2022). While the effect-size of the expression of a single gene in microbial organisms is likely to be considerably less, due to the ability of pathogens to compensate for gene loss by upregulating alternative pathways, targeting multiple genes may be a plausible approach for microorganisms. As sequencing technologies including single cell transcriptomics continue to advance, and the cost associated with them reduces, we may be able to design more elaborate experiments that could unveil further insights in the regulation of bacterial gene expression. The nasopharynx which is the natural niche of *S. pneumoniae*, is resident to multiple other bacterial and viral agents which compete with *S. pneumoniae* for nutrition and resources (Watkins *et al.* 2022; Man, de Steenhuijsen Piters, and Bogaert 2017). The dynamic of microbial species in the nasopharyngeal microbiome ecosystem is impacted by external factors including vaccines, antibiotics, and nutrition (Watkins *et al.* 2022; Man, de Steenhuijsen Piters, and Bogaert 2017). In addition, viral pathogens such as influenza virus have been known to be associated with pneumococcal carriage and disease (Watkins *et*

*al.* 2022; Man, de Steenhuijsen Piters, and Bogaert 2017). Therefore, the ability to simultaneously investigate the transcriptome of multiple bacterial and viral agents, which is not possible with current technologies, will refine our understanding of the interplay in pneumococcal adaptation in the nasopharyngeal environment. *In vitro* dual host-pathogen RNA-Seq using primary cells lines is currently the best available method for simultaneously interrogating the transcriptome of a single bacteria strain and the host (Westermann, Gorski, and Vogel 2012; Aprianto *et al.* 2016; Westermann *et al.* 2016; Aulicino *et al.* 2022; Lopez-Agudelo *et al.* 2022). Results in this thesis have emphasised the role pneumococcal metabolism plays in bacterial adaptation and virulence. Thus, for the true understanding of the global adaptation of *S. pneumoniae*, integrating multi-omics studies which simultaneously characterise the pathogen at multiple levels including transcriptome, proteome, and metabolome will be necessary to have a fuller picture of molecular interactions. We expect that such studies will complement the work presented in this thesis and uncover novel insights into pathogenic bacterial behaviour which may accelerate the development of more effective pneumococcal interventions.

### 9.5 Closing remarks

Different serotypes of *S. pneumoniae* differ in propensity to colonise the human nasopharynx and to cause invasive disease. The mechanisms that govern these differences are not fully understood. RNA-Seq was used in this thesis to evaluate the genome-wide transcriptomic response of clinical pneumococcal strains in a set of controlled conditions, to elucidate how *S. pneumoniae* fine-tune's gene expression during adaptation to specific environments. This work represents the first comparative transcriptomic dataset of a highly invasive (serotype 1) and a relatively less invasive (serotype 6B) pneumococcal strains. The results suggests that differences in behaviour of pneumococcal strains is likely associated with differences in the regulation and expression of core pneumococcal genes, providing interesting avenues for further hypothesis testing which could potentially lead to the development of novel interventions for controlling pneumococcal disease.

## Chapter 10 : Appendices

## 10.1 Appendix 1: Front page of publication resulting from work in Chapter 4.



GENOME SEQUENCES



# Complete Genome Sequence of *Streptococcus pneumoniae* Strain BVJ1JL, a Serotype 1 Carriage Isolate from Malawi

Modupeh Betts,<sup>a</sup> Seth Jarvis,<sup>b</sup> Aaron Jeffries,<sup>c</sup> Andrea Gori,<sup>a</sup> Chrispin Chaguz,<sup>d,e,f</sup> Jacqueline Msefula,<sup>g</sup> Caroline M. Weight,<sup>a</sup> Brenda Kwambana-Adams,<sup>a</sup> Neil French,<sup>f,g</sup> Todd D. Swarthout,<sup>a,g</sup> Jeremy S. Brown,<sup>h</sup> Robert S. Heyderman<sup>a,g</sup>

<sup>a</sup>NIHR Global Health Research Unit on Mucosal Pathogens, Division of Infection and Immunity, University College London, London, United Kingdom

<sup>b</sup>UCL Genetics Institute, University College London, London, United Kingdom

<sup>c</sup>University of Exeter, Exeter, United Kingdom

<sup>d</sup>Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, United Kingdom

<sup>e</sup>Darwin College, University of Cambridge, Cambridge, United Kingdom

<sup>f</sup>Department of Clinical Infection, Microbiology and Immunology, Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Liverpool, United Kingdom

<sup>g</sup>Malawi-Liverpool-Wellcome Trust Clinical Research Programme Blantyre, Blantyre, Malawi

<sup>h</sup>Department of Respiratory Medicine, Centre for Inflammation and Tissue Repair, University College London, London, United Kingdom

**ABSTRACT** *Streptococcus pneumoniae* is a leading cause of pneumonia, meningitis, and bacteremia. Serotype 1 is rarely carried but is commonly associated with invasive pneumococcal disease, and in the African “meningitis belt,” it is prone to cause cyclical epidemics. We report the complete genome sequence of *S. pneumoniae* serotype 1 strain BVJ1JL, isolated in Malawi.

*Streptococcus pneumoniae*, a Gram-positive bacterium, is a leading cause of childhood mortality worldwide (1). At least 100 different capsular serotypes of *S. pneumoniae* have been described (2). Serotype 1 is among the most commonly isolated serotypes from blood or cerebrospinal fluid (CSF) (3, 4).

Strain BVJ1JL was isolated in 2015 from a nasopharyngeal swab (NPS) obtained from a 9-year-old child in Blantyre, Malawi (5). The study protocol was approved by the College of Medicine Research and Ethics Committee, University of Malawi (P.02/15/1677), and the Liverpool School of Tropical Medicine Research Ethics Committee, UK (14.056). The primary NPS, retrieved from storage in skim milk-tryptone-glucose-glycerol (STGG) medium, was plated onto Columbia blood agar (CBA) supplemented with 5% horse blood and incubated overnight at 37°C and 5% CO<sub>2</sub>. A single colony was then picked and purified on a fresh CBA plate. *S. pneumoniae* was confirmed by morphology, optochin test, and Gram stain. The capsular type was assessed using a serological latex agglutination test kit (ImmuLex Pneumotest; SSI Diagnostica) and confirmed genomically using PneumoCat (6). DNA was isolated from lawn plate cultures of frozen stocks incubated overnight at 37°C and 5% CO<sub>2</sub> on CBA. The Qiagen Genomic-tip 500/G DNA kit was used to isolate DNA for PacBio sequencing, following the manufacturer's protocol. Lysis buffer was supplemented with 30 mg/ml lysozyme (Sigma-Aldrich) and 50 units mutanolysin (Sigma-Aldrich). DNA was sheared using a g-TUBE device (Covaris) with a target length of 10 kb, and library preparation was performed according to the protocol “Preparing Multiplexed Microbial Libraries Using SMRTbell Express Template Prep Kit 2.0,” with the Barcoded Overhang Adapter Kit-8A (Pacific Biosciences) at the University of Exeter, UK. The pooled samples were purified and size selected to remove SMRTbell templates of >3 kb using AMPure PB beads (Pacific Biosciences). A 10-h capture using a 1M single-molecule real-time (SMRT) cell was performed on a PacBio Sequel instrument using Sequel v3 chemistry (7). In total, 295,697 reads were generated for BVJ1JL with a read N<sub>50</sub> length of

**Citation** Betts M, Jarvis S, Jeffries A, Gori A, Chaguz C, Msefula J, Weight CM, Kwambana-Adams B, French N, Swarthout TD, Brown JS, Heyderman RS. 2021. Complete genome sequence of *Streptococcus pneumoniae* strain BVJ1JL, a serotype 1 carriage isolate from Malawi. Microbiol Resour Announc 10:e00715-21. <https://doi.org/10.1128/MRA.00715-21>.

**Editor** Steven R. Gill, University of Rochester School of Medicine and Dentistry

**Copyright** © 2021 Betts et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Modupeh Betts, modupeh.betts.17@ucl.ac.uk.

**Received** 14 July 2021

**Accepted** 4 September 2021

**Published** 30 September 2021

**10.2 Appendix 2: Detailed preparation of 7H9+ growth medium**

To prepare 50mM potassium phosphate buffer (pH 7) (to dissolve catalase in):

Add 135.5mg  $\text{KH}_2\text{PO}_4$  + 261.5mg  $\text{K}_2\text{HPO}_4$  in 50ml  $\text{dH}_2\text{O}$

To prepare 20% tyloxapol:

Add 5ml 100% Tyloxapol + 20ml  $\text{dH}_2\text{O}$  (50ml Falcon tube);

80°C heat block till dissolved.

Store at 4°C wrapped in foil – good for up to 6 months

**Appendix Table 2A:** Recipe for Solution IV

Reagent	In 1L	100x (50ml)
Biotin	0.015mg	0.075mg
Choline chloride	5 mg	25mg
Nicotinamide	0.6 mg	3mg
Ca pantothenate	2.4mg	12mg
Pyridoxal HCl	0.6mg	3mg
(-) Riboflavin	0.3mg	1.5mg
Thiamine.HCl	0.6mg	3mg
L-cysteine.HCl	100mg	500mg
L-glutamine	20mg	100mg
Sodium pyruvate	0.8g	4g

Dissolve the amounts in **Appendix Table 2A** in 50ml  $\text{dH}_2\text{O}$ .

Filter-sterilise (0.2um filter), distribute in 1ml aliquots, store at -80°C – good for ~1 year.

**Recipe for Solution II:**

Add 2.35g 7H9 broth base (Sigma) to 500mL  $\text{dH}_2\text{O}$

Dissolve the amino acids in **Appendix Table 2B** in 500mL 7H9 broth base solution:

**Appendix Table 2B:** Amino acids

L-arg	100mg	L-leu	3.3mg
Glycine	60mg	L-ile	3.3mg
L-his	75mg	L-val	2.9mg
L-met	90mg	Uracil	0.5mg
L-thr	87.5mg	L-lys	210mg
L-asn	5mg		

Store at 4°C – Solution II is good for maximum 2 weeks.

**Recipe for final 7H9<sup>+</sup> media:**

Make 10% BSA – 1g BSA in 10ml dH<sub>2</sub>O; Mix on roller at 4°C (cold room) – prepare up to 24 hours before use.

Make fresh 0.1% catalase – 1mg/ml in 50mM potassium phosphate buffer (pH 7.0) – always make fresh as catalase degrades once in solution.

To 100ml of Solution II, add:

5ml of 10% BSA

250ul of 20% Tyloxapol

300ul of 0.1% catalase

Filter-sterilise the entire solution. Distribute media into 50ml Falcons and supplement with appropriate volumes of sterile sugar stocks (glucose, galactose, sucrose, etc) – Note that THY contains 0.2% or ~11mM of glucose and this is the amount Dr. Mane uses (i.e., 800ul of 25% glucose solution per 100ml of media).

Inoculate with pre-determined amount of frozen pneumococcal stock.

Incubate at 37°C/5% CO<sub>2</sub> with loosened lids on the Falcon tubes.

**10.3 Appendix 3: Detailed preparation of NMC growth medium (Aprianto *et al.* 2016)****Appendix Table 3A: Amino acids**

Amino Acids	(mM)	Amount in 500mL (g)	Cat #	Company
L-alanine	2.69	1.5	05130-25G	Sigma
		0.78	11009-25G-	Sigma
L-arginine	0.71		F	
L-asparagine	2.34	2.2	A4159-25G	Sigma
L-aspartate	3	2.6	11189-100G	Sigma
L-cysteine	2.538	1.5	C7352-25G	Sigma
L-glutamate	2.46	3.1	49449-100G	Sigma
L-glutamine	2.68	2.45	49419-25G	Sigma
Glycine	2.34	1.1	50046-50G	Sigma
L-histidine	0.72	0.95	H8000-5G	Sigma
L-isoleucine	1.62	1.33	I2752-10G	Sigma
L-leucine	3.48	2.85	61819-25G	Sigma
L-lysine	2.41	2.75	L5501-25G	Sigma
L-methionine	0.83	0.78	M9625-25G	Sigma
L-phenylalanine	1.67	1.73	78019-25G	Sigma
Proline	5.87	4.23	P5607-25G	Sigma
L-serine	3.23	2.13	S4311-25G	Sigma
L-threonine	1.88	2.4	89179-10G	Sigma
L-tryptophan	0.25	0.32	93659-10G	Sigma
L-valine	2.77	2.0	V0500-25G	Sigma

*Adjust pH to 6.5, filtered (0.22µm) and store at -20 °C in aluminium foil.*

**Appendix Table 3B: Basal Nose Solution**



Basal Nose Solution	(mM)	Amount in 500mL	Cat #	Company
Cysteine-HCL		250mg	30120-10G	Sigma
Dipotassium phosphate	14.3	9mL		Sigma
Monopotassium phosphate	22	14mL		Sigma
		375mg	25102-500G-	Sigma
Ammonium citrate	2.65		R	
Sodium acetate	12.19	625mg	58750-500G	Sigma

*Make 1N of  $K_2HPO_4$  and  $KH_2PO_4$  solutions and added the respective volumes. Keep solution at 4 °C for long term use. Filter final solution (0.22 $\mu$ m) and store at 4C.*

**Appendix Table 3C: Micronutrients**

Micronutrients	(mM)	Amount in 500mL (g)	Cat #	Company
		10	M8266-	Sigma
Magnesium chloride	2.1		100G	
Calcium chloride	0.34	1.9	C1016-500G	Sigma
		250mg	83265-	Sigma
Zinc sulfate ( $\times 10^{-3}$ )*	30.97		250ML	

*Solution to be filtered (0.22 $\mu$ m) and kept at 4 °C.*

*\*Reagent ordered as a 2M solution.*

**Appendix Table 3D: Nucleobases**

<b>Nucleobases</b>	<b>(<math>\mu</math>M)</b>	<b>Amount in 500mL (g)</b>	<b>Cat #</b>	<b>Company</b>
Adenine	74	0.5	A2786-5G	Sigma
Guanine	66.17	0.5	U1128-25G	Sigma
Uracil	89.22	0.5	X7375-10G	Sigma
		0.5	G11950-	Sigma
Xanthine	65.74		25G	

*Dissolve in 0.1N NaOH, filter (0.22 $\mu$ m), and store at 4 °C*

**Appendix Table 3E: Vitamins**

<b>Vitamins</b>	<b>(<math>\mu</math>M)</b>	<b>Amount in 500mL (mg)</b>	<b>Cat #</b>	<b>Company</b>
B1, Thiamine-HCl	2.965	50	T1270-25G	Sigma
B2, Riboflavin	2.657	50	R9504-25G	Sigma
B3, Nicotinic acid, niacinamide	8.123	50	72345-50G	Sigma
		50	P5155-	Sigma
B5, Calcium pantothenate	2.098		100G	
B6, Pyridoxamine-HCl*	12.216	80	P9380-1G	Sigma
B6, Pyridoxine-HCl	9.726	125	P6280-10G	Sigma
B7, D-Biotin	10.233	125	BP232-1	Fisher
B9, Folic acid	2.266	50	20515	Cayman
		50	V2876-	Sigma
B12, Cyanocobalamin	0.738		500G	
Lipoic acid	7.27	75	T1395-1G	Sigma
		250	A6928-	Sigma
Sodium-p-aminobenzoate	31.423		100G	

*Riboflavin to be dissolved in 350mL mQ water and heated up to 70 °C. Other vitamins to be added when solution is around 30 °C. Biotin and folic acid to be dissolved in 2N NaOH, added to the solution to make 500mL and pH adjusted to 6.5, filtered (0.22 $\mu$ m), and kept at 4 °C.*

**Appendix Table 3E: Others**

			<b>Cat#</b>	<b>Company</b>
<b>Miscellaneous</b>	<b>Concentration</b>	<b>For media</b>		
Choline hydrochloride ( $\mu\text{M}$ )	72	2.5g/L	C7527-100G	Sigma
N-Acetylglucosamine	5.79	20%	PHR1432-1G	Sigma
Serum Albumin ( $\text{g.L}^{-1}$ )	1	25%	A7030-10G	Sigma
Sodium pyruvate (mM)	1.1	10%	P2256-25G	Sigma

*Prepare each separately, Filter (0.22 $\mu\text{m}$ ), and store at 4 °C*

#### **Preparation of nose mimicking condition (NMC) media**

To prepare 100mL of “Nose mimicking condition” media, sterilely mix 80mL of nose basal solution, 8mL of amino acids, 1mL of micronutrients, 1mL of nucleobases, 1mL of vitamins, 100uL of pyruvate, 0.8mL of choline, 0.4mL of 25% BSA and 0.64mL of N-Acetylglucosamine. pH adjusts to 7.0 and add mQ to 100mL. Filter

#### **Preparation of lung mimicking condition (LMC) media**

100 ml lung-like medium was made by sterilely mixing 80 mL nose basal solution, 8 mL amino acids, 1 mL micronutrients, 1 mL nucleobases, 1 mL vitamins, 0.1 mL pyruvate, 0.8 ml choline, 1.2 mL 25% BSA and 0.32 mL glucosamine solutions. pH was adjusted to 7.0 and sterile mQ was added up to 100 ml

# Reference list

Abu-Jamous, B., and S. Kelly. 2018. 'Clust: automatic extraction of optimal co-expressed gene clusters from gene expression data', *Genome Biol*, 19: 172.

Adamou, J. E., J. H. Heinrichs, A. L. Erwin, W. Walsh, T. Gayle, M. Dormitzer, R. Dagan, Y. A. Brewah, P. Barren, R. Lathigra, S. Langermann, S. Koenig, and S. Johnson. 2001. 'Identification and characterization of a novel family of pneumococcal proteins that are protective against sepsis', *Infect Immun*, 69: 949-58.

Afzal, M., S. Shafeeq, and O. P. Kuipers. 2016. 'Methionine-mediated gene expression and characterization of the CmhR regulon in *Streptococcus pneumoniae*', *Microb Genom*, 2: e000091.

Aguilera, E. R., and L. L. Lenz. 2020. 'Inflammation as a Modulator of Host Susceptibility to Pulmonary Influenza, Pneumococcal, and Co-Infections', *Front Immunol*, 11: 105.

Akhter, F., E. Womack, J. E. Vidal, Y. Le Breton, K. S. McIver, S. Pawar, and Z. Eichenbaum. 2020. 'Hemoglobin stimulates vigorous growth of *Streptococcus pneumoniae* and shapes the pathogen's global transcriptome', *Sci Rep*, 10: 15202.

Ali, M. Q., T. P. Kohler, G. Burchhardt, A. Wust, N. Henck, R. Bolsmann, F. Voss, and S. Hammerschmidt. 2020. 'Extracellular Pneumococcal Serine Proteases Affect Nasopharyngeal Colonization', *Front Cell Infect Microbiol*, 10: 613467.

Ali, M. Q., T. P. Kohler, L. Schulig, G. Burchhardt, and S. Hammerschmidt. 2021. 'Pneumococcal Extracellular Serine Proteases: Molecular Analysis and Impact on Colonization and Disease', *Front Cell Infect Microbiol*, 11: 763152.

- Alloing, G., B. Martin, C. Granadel, and J. P. Claverys. 1998. 'Development of competence in *Streptococcus pneumoniae*: pheromone autoinduction and control of quorum sensing by the oligopeptide permease', *Mol Microbiol*, 29: 75-83.
- Almeida, S. T., T. Pedro, A. C. Paulo, H. de Lencastre, and R. Sa-Leao. 2020. 'Re-evaluation of *Streptococcus pneumoniae* carriage in Portuguese elderly by qPCR increases carriage estimates and unveils an expanded pool of serotypes', *Sci Rep*, 10: 8373.
- Andreini, C., I. Bertini, G. Cavallaro, G. L. Holliday, and J. M. Thornton. 2008. 'Metal ions in biological catalysis: from enzyme databases to general principles', *J Biol Inorg Chem*, 13: 1205-18.
- Andrews, Simon C., Andrea K. Robinson, and Francisco Rodríguez-Quñones. 2003. 'Bacterial iron homeostasis', *FEMS Microbiology Reviews*, 27: 215-37.
- Aprianto, R., J. Slager, S. Holsappel, and J. W. Veening. 2016. 'Time-resolved dual RNA-seq reveals extensive rewiring of lung epithelial and pneumococcal transcriptomes during early infection', *Genome Biol*, 17: 198.
- Aprianto, Rieza, Jelle Slager, Siger Holsappel, and Jan-Willem Veening. 2018. 'High-resolution analysis of the pneumococcal transcriptome under a wide range of infection-relevant conditions', *Nucleic Acids Research*.
- Aulicino, A., A. Antanaviciute, J. Frost, A. Sousa Geros, E. Mellado, M. Attar, M. Jagielowicz, P. Hublitz, J. Sinz, L. Preciado-Llanes, G. Napolitani, R. Bowden, H. Koohy, H. Drakesmith, and A. Simmons. 2022. 'Dual RNA sequencing reveals dendritic cell reprogramming in response to typhoidal *Salmonella* invasion', *Commun Biol*, 5: 111.
- Avery, O. T., C. M. Macleod, and M. McCarty. 1944. 'Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from *Pneumococcus* Type lii', *J Exp Med*, 79: 137-58.

Ayoola, M. B., L. A. Shack, M. F. Nakamya, J. A. Thornton, E. Swiatlo, and B. Nanduri. 2019. 'Polyamine Synthesis Effects Capsule Expression by Reduction of Precursors in *Streptococcus pneumoniae*', *Front Microbiol*, 10: 1996.

Balsells, E., R. Dagan, I. Yildirim, P. P. Gounder, A. Steens, C. Munoz-Almagro, C. Mameli, R. Kandasamy, N. Givon Lavi, L. Daprai, A. van der Ende, K. Trzcinski, S. A. Nzenze, S. Meiring, D. Foster, L. R. Bulkow, K. Rudolph, A. Valero-Rello, S. Ducker, D. F. Vestrheim, A. von Gottberg, S. I. Pelton, G. Zuccotti, A. J. Pollard, E. A. M. Sanders, H. Campbell, S. A. Madhi, H. Nair, and M. H. Kyaw. 2018. 'The relative invasive disease potential of *Streptococcus pneumoniae* among children after PCV introduction: A systematic review and meta-analysis', *J Infect*, 77: 368-78.

Bar, A., L. Argaman, Y. Altuvia, and H. Margalit. 2021. 'Prediction of Novel Bacterial Small RNAs From RIL-Seq RNA-RNA Interaction Data', *Front Microbiol*, 12: 635070.

Barczak, A. K., J. E. Gomez, B. B. Kaufmann, E. R. Hinson, L. Cosimi, M. L. Borowsky, A. B. Onderdonk, S. A. Stanley, D. Kaur, K. F. Bryant, D. M. Knipe, A. Sloutsky, and D. T. Hung. 2012. 'RNA signatures allow rapid identification of pathogens and antibiotic susceptibilities', *Proc Natl Acad Sci U S A*, 109: 6217-22.

Bashir, A., A. Klammer, W. P. Robins, C. S. Chin, D. Webster, E. Paxinos, D. Hsu, M. Ashby, S. Wang, P. Peluso, R. Sebra, J. Sorenson, J. Bullard, J. Yen, M. Valdovino, E. Mollova, K. Luong, S. Lin, B. LaMay, A. Joshi, L. Rowe, M. Frace, C. L. Tarr, M. Turnsek, B. M. Davis, A. Kasarskis, J. J. Mekalanos, M. K. Waldor, and E. E. Schadt. 2012. 'A hybrid approach for the automated finishing of bacterial genomes', *Nat Biotechnol*, 30: 701-07.

Basset, A., E. Wall, E. Mitsi, C. Deshusses, R. Daly, S. Pojar, J. Reine, J. A. Guerra-Assuncao, B. Denis, S. P. Jochems, R. Heyderman, J. Brown, Y. J. Lu, D. M. Ferreira, and R. Malley. 2022. 'Targeted Transcriptomic Screen of Pneumococcal Genes Expressed during Murine and Human Infection', *Infect Immun*: e0017522.

Belitsky, B. R., J. Brill, E. Bremer, and A. L. Sonenshein. 2001. 'Multiple genes for the last step of proline biosynthesis in *Bacillus subtilis*', *J Bacteriol*, 183: 4389-92.

Bentley, D. R., S. Balasubramanian, H. P. Swerdlow, G. P. Smith, J. Milton, C. G. Brown, K. P. Hall, D. J. Evers, C. L. Barnes, H. R. Bignell, J. M. Boutell, J. Bryant, R. J. Carter, R. Keira Cheetham, A. J. Cox, D. J. Ellis, M. R. Flatbush, N. A. Gormley, S. J. Humphray, L. J. Irving, M. S. Karbelashvili, S. M. Kirk, H. Li, X. Liu, K. S. Maisinger, L. J. Murray, B. Obradovic, T. Ost, M. L. Parkinson, M. R. Pratt, I. M. Rasolonjatovo, M. T. Reed, R. Rigatti, C. Rodighiero, M. T. Ross, A. Sabot, S. V. Sankar, A. Scally, G. P. Schroth, M. E. Smith, V. P. Smith, A. Spiridou, P. E. Torrance, S. S. Tzonev, E. H. Vermaas, K. Walter, X. Wu, L. Zhang, M. D. Alam, C. Anastasi, I. C. Aniebo, D. M. Bailey, I. R. Bancarz, S. Banerjee, S. G. Barbour, P. A. Baybayan, V. A. Benoit, K. F. Benson, C. Bevis, P. J. Black, A. Boodhun, J. S. Brennan, J. A. Bridgham, R. C. Brown, A. A. Brown, D. H. Buermann, A. A. Bundu, J. C. Burrows, N. P. Carter, N. Castillo, E. Catenazzi M. Chiara, S. Chang, R. Neil Cooley, N. R. Crake, O. O. Dada, K. D. Diakoumakos, B. Dominguez-Fernandez, D. J. Earnshaw, U. C. Egbujor, D. W. Elmore, S. S. Etchin, M. R. Ewan, M. Fedurco, L. J. Fraser, K. V. Fuentes Fajardo, W. Scott Furey, D. George, K. J. Gietzen, C. P. Goddard, G. S. Golda, P. A. Granieri, D. E. Green, D. L. Gustafson, N. F. Hansen, K. Harnish, C. D. Haudenschild, N. I. Heyer, M. M. Hims, J. T. Ho, A. M. Horgan, K. Hoschler, S. Hurwitz, D. V. Ivanov, M. Q. Johnson, T. James, T. A. Huw Jones, G. D. Kang, T. H. Kerelska, A. D. Kersey, I. Khrebtukova, A. P. Kindwall, Z. Kingsbury, P. I. Kokko-Gonzales, A. Kumar, M. A. Laurent, C. T. Lawley, S. E. Lee, X. Lee, A. K. Liao, J. A. Loch, M. Lok, S. Luo, R. M. Mammen, J. W. Martin, P. G. McCauley, P. McNitt, P. Mehta, K. W. Moon, J. W. Mullens, T. Newington, Z. Ning, B. Ling Ng, S. M. Novo, M. J. O'Neill, M. A. Osborne, A. Osnowski, O. Ostadan, L. L. Paraschos, L. Pickering, A. C. Pike, A. C. Pike, D. Chris Pinkard, D. P. Pliskin, J. Podhasky, V. J. Quijano, C. Raczy, V. H. Rae, S. R. Rawlings, A. Chiva Rodriguez, P. M. Roe, J. Rogers, M. C. Rogert Bacigalupo, N. Romanov, A. Romieu, R. K. Roth, N. J. Rourke, S. T. Ruediger, E. Rusman, R. M. Sanches-Kuiper, M. R. Schenker, J. M. Seoane, R. J. Shaw, M. K. Shiver, S. W. Short, N. L. Sizto, J. P. Sluis, M. A. Smith, J. Ernest Sohna Sohna, E. J. Spence, K. Stevens, N. Sutton, L. Szajkowski, C. L. Tregidgo, G. Turcatti, S. Vandevondele, Y. Verhovsky, S. M. Virk, S. Wakelin, G. C. Walcott, J. Wang, G. J. Worsley, J. Yan, L. Yau, M. Zuerlein, J. Rogers, J. C. Mullikin, M. E. Hurles, N. J. McCooke, J. S. West, F. L. Oaks, P. L. Lundberg, D. Klenerman, R. Durbin, and A. J. Smith. 2008. 'Accurate whole human genome sequencing using reversible terminator chemistry', *Nature*, 456: 53-9.

Berry, A. M., R. A. Lock, D. Hansman, and J. C. Paton. 1989. 'Contribution of autolysin to virulence of *Streptococcus pneumoniae*', *Infect Immun*, 57: 2324-30.

Berry, A. M., and J. C. Paton. 1996. 'Sequence heterogeneity of PsaA, a 37-kilodalton putative adhesin essential for virulence of *Streptococcus pneumoniae*', *Infect Immun*, 64: 5255-62.

Berry, A. M., J. Yother, D. E. Briles, D. Hansman, and J. C. Paton. 1989. 'Reduced virulence of a defined pneumolysin-negative mutant of *Streptococcus pneumoniae*', *Infect Immun*, 57: 2037-42.

Betin, V., C. Penaranda, N. Bandyopadhyay, R. Yang, A. Abitua, R. P. Bhattacharyya, A. Fan, R. Avraham, J. Livny, N. Shores, and D. T. Hung. 2019. 'Hybridization-based capture of pathogen mRNA enables paired host-pathogen transcriptional analysis', *Sci Rep*, 9: 19244.

Betts, M., S. Jarvis, A. Jeffries, A. Gori, C. Chaguz, J. Msefula, C. M. Weight, B. Kwambana-Adams, N. French, T. D. Swarthout, J. S. Brown, and R. S. Heyderman. 2021. 'Complete Genome Sequence of *Streptococcus pneumoniae* Strain BVJ1JL, a Serotype 1 Carriage Isolate from Malawi', *Microbiol Resour Announc*, 10: e0071521.

Bhattacharya, A., A. M. Hamilton, H. Furberg, E. Pietzak, M. P. Purdue, M. A. Troester, K. A. Hoadley, and M. I. Love. 2021. 'An approach for normalization and quality control for NanoString RNA expression data', *Brief Bioinform*, 22.

Bidossi, A., L. Mulas, F. Decorosi, L. Colomba, S. Ricci, G. Pozzi, J. Deutscher, C. Viti, and M. R. Oggioni. 2012. 'A functional genomics approach to establish the complement of carbohydrate transporters in *Streptococcus pneumoniae*', *PLoS One*, 7: e33320.

Bogaert, D., R. de Groot, and P. W. M. Hermans. 2004. 'Streptococcus pneumoniae colonisation: the key to pneumococcal disease', *The Lancet Infectious Diseases*, 4: 144-54.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. 'Trimmomatic: a flexible trimmer for Illumina sequence data', *Bioinformatics*, 30: 2114-20.



- Bricio-Moreno, L., C. Chaguza, R. Yahya, R. K. Shears, J. E. Cornick, K. Hokamp, M. Yang, D. R. Neill, N. French, J. C. D. Hinton, D. B. Everett, and A. Kadioglu. 2020. 'Lower Density and Shorter Duration of Nasopharyngeal Carriage by Pneumococcal Serotype 1 (ST217) May Explain Its Increased Invasiveness over Other Serotypes', *mBio*, 11.
- Briles, D. E., E. Ades, J. C. Paton, J. S. Sampson, G. M. Carlone, R. C. Huebner, A. Virolainen, E. Swiatlo, and S. K. Hollingshead. 2000. 'Intranasal immunization of mice with a mixture of the pneumococcal proteins PsaA and PspA is highly protective against nasopharyngeal carriage of *Streptococcus pneumoniae*', *Infect Immun*, 68: 796-800.
- Briles, D. E., M. J. Crain, B. M. Gray, C. Forman, and J. Yother. 1992. 'Strong association between capsular type and virulence for mice among human isolates of *Streptococcus pneumoniae*', *Infect Immun*, 60: 111-6.
- Brinster, S., G. Lamberet, B. Staels, P. Trieu-Cuot, A. Gruss, and C. Poyart. 2009. 'Type II fatty acid synthesis is not a suitable antibiotic target for Gram-positive pathogens', *Nature*, 458: 83-6.
- Brondsted, L., M. T. Andersen, M. Parker, K. Jorgensen, and H. Ingmer. 2005. 'The HtrA protease of *Campylobacter jejuni* is required for heat and oxygen tolerance and for optimal interaction with human epithelial cells', *Appl Environ Microbiol*, 71: 3205-12.
- Brooks, L. R. K., and G. I. Mias. 2018. '*Streptococcus pneumoniae*'s Virulence and Host Immunity: Aging, Diagnostics, and Prevention', *Front Immunol*, 9: 1366.
- Browall, S., M. Norman, J. Tangrot, I. Galanis, K. Sjostrom, J. Dagerhamn, C. Hellberg, A. Pathak, T. Spadafina, A. Sandgren, P. Battig, O. Franzen, B. Andersson, A. Ortqvist, S. Normark, and B. Henriques-Normark. 2014. 'Intraclonal variations among *Streptococcus pneumoniae* isolates influence the likelihood of invasive disease in children', *J Infect Dis*, 209: 377-88.
- Brown, A. O., B. Mann, G. Gao, J. S. Hankins, J. Humann, J. Giardina, P. Faverio, M. I. Restrepo, G. V. Halade, E. M. Mortensen, M. L. Lindsey, M. Hanes, K. I. Happel, S. Nelson,

G. J. Bagby, J. A. Lorent, P. Cardinal, R. Granados, A. Esteban, C. J. LeSaux, E. I. Tuomanen, and C. J. Orihuela. 2014. 'Streptococcus pneumoniae translocates into the myocardium and forms unique microlesions that disrupt cardiac function', *PLoS Pathog*, 10: e1004383.

Brown, J. S., S. M. Gilliland, and D. W. Holden. 2001. 'A Streptococcus pneumoniae pathogenicity island encoding an ABC transporter involved in iron uptake and virulence', *Mol Microbiol*, 40: 572-85.

Brown, J. S., S. M. Gilliland, J. Ruiz-Albert, and D. W. Holden. 2002. 'Characterization of pit, a Streptococcus pneumoniae iron uptake ABC transporter', *Infect Immun*, 70: 4389-98.

Brown, J. S., A. D. Ogunniyi, M. C. Woodrow, D. W. Holden, and J. C. Paton. 2001. 'Immunization with components of two iron uptake ABC transporters protects mice against systemic Streptococcus pneumoniae infection', *Infect Immun*, 69: 6702-6.

Brown, Jeremy S., and David W. Holden. 2002. 'Iron acquisition by Gram-positive bacterial pathogens', *Microbes and Infection*, 4: 1149-56.

Brown, L. R., S. M. Gunnell, A. N. Cassella, L. E. Keller, L. A. Scherkenbach, B. Mann, M. W. Brown, R. Hill, N. C. Fitzkee, J. W. Rosch, E. I. Tuomanen, and J. A. Thornton. 2016. 'AdcAll of Streptococcus pneumoniae Affects Pneumococcal Invasiveness', *PLoS One*, 11: e0146785.

Bruckner, R., M. Nuhn, P. Reichmann, B. Weber, and R. Hakenbeck. 2004. 'Mosaic genes and mosaic chromosomes-genomic variation in Streptococcus pneumoniae', *Int J Med Microbiol*, 294: 157-68.

Brueggemann, A. B., T. E. Peto, D. W. Crook, J. C. Butler, K. G. Kristinsson, and B. G. Spratt. 2004. 'Temporal and geographic stability of the serogroup-specific invasive disease potential of Streptococcus pneumoniae in children', *J Infect Dis*, 190: 1203-11.

Buckwalter, C. M., and S. J. King. 2012. 'Pneumococcal carbohydrate transport: food for thought', *Trends Microbiol*, 20: 517-22.

- Caglar, M. U., J. R. Houser, C. S. Barnhart, D. R. Boutz, S. M. Carroll, A. Dasgupta, W. F. Lenoir, B. L. Smith, V. Sridhara, D. K. Sydykova, D. Vander Wood, C. J. Marx, E. M. Marcotte, J. E. Barrick, and C. O. Wilke. 2017. 'The E. coli molecular phenotype under different growth conditions', *Sci Rep*, 7: 45303.
- Call, D. R., M. K. Bakko, M. J. Krug, and M. C. Roberts. 2003. 'Identifying antimicrobial resistance genes with DNA microarrays', *Antimicrob Agents Chemother*, 47: 3290-5.
- Campbell, E. A., S. Y. Choi, and H. R. Masure. 1998. 'A competence regulon in *Streptococcus pneumoniae* revealed by genomic analysis', *Mol Microbiol*, 27: 929-39.
- Canals, R., D. L. Hammarlof, C. Kroger, S. V. Owen, W. Y. Fong, L. Lacharme-Lora, X. Zhu, N. Wenner, S. E. Carden, J. Honeycutt, D. M. Monack, R. A. Kingsley, P. Brownridge, R. R. Chaudhuri, W. P. M. Rowe, A. V. Predeus, K. Hokamp, M. A. Gordon, and J. C. D. Hinton. 2019. 'Adding function to the genome of African *Salmonella* Typhimurium ST313 strain D23580', *PLoS Biol*, 17: e3000059.
- Canvin, J. R., A. P. Marvin, M. Sivakumaran, J. C. Paton, G. J. Boulnois, P. W. Andrew, and T. J. Mitchell. 1995. 'The role of pneumolysin and autolysin in the pathology of pneumonia and septicemia in mice infected with a type 2 pneumococcus', *J Infect Dis*, 172: 119-23.
- Cao, K., F. Lai, X. L. Zhao, Q. X. Wei, X. Y. Miao, R. Ge, Q. Y. He, and X. Sun. 2018. 'The mechanism of iron-compensation for manganese deficiency of *Streptococcus pneumoniae*', *J Proteomics*, 184: 62-70.
- Capes-Davis, A., A. Bairoch, T. Barrett, E. C. Burnett, W. G. Dirks, E. M. Hall, L. Healy, D. A. Kniss, C. Korch, Y. Liu, R. M. Neve, R. W. Nims, B. Parodi, R. E. Schweppe, D. R. Storts, and F. Tian. 2019. 'Cell Lines as Biological Models: Practical Steps for More Reliable Research', *Chem Res Toxicol*, 32: 1733-36.
- Carreno, D., J. J. Wanford, Z. Jasiunaite, R. G. Hames, W. Y. Chung, A. R. Dennison, K. Straatman, L. Martinez-Pomares, M. Pareek, C. J. Orihuela, M. I. Restrepo, W. S. Lim, P. W.

- Andrew, E. R. Moxon, and M. R. Oggioni. 2021. 'Splenic macrophages as the source of bacteraemia during pneumococcal pneumonia', *EBioMedicine*, 72: 103601.
- Carvalho Mda, G., M. L. Tondella, K. McCaustland, L. Weidlich, L. McGee, L. W. Mayer, A. Steigerwalt, M. Whaley, R. R. Facklam, B. Fields, G. Carlone, E. W. Ades, R. Dagan, and J. S. Sampson. 2007. 'Evaluation and improvement of real-time PCR assays targeting *lytA*, *ply*, and *psaA* genes for detection of pneumococcal DNA', *J Clin Microbiol*, 45: 2460-6.
- Cassat, J. E., and E. P. Skaar. 2013. 'Iron in infection and immunity', *Cell Host Microbe*, 13: 509-19.
- Chaguza, C., J. E. Cornick, S. R. Harris, C. P. Andam, L. Bricio-Moreno, M. Yang, F. Yalcin, S. Ousmane, S. Govindpersad, M. Senghore, C. Ebruke, M. Du Plessis, A. M. Kiran, G. Pluschke, B. Sigauque, L. McGee, K. P. Klugman, P. Turner, J. Corander, J. Parkhill, J. M. Collard, M. Antonio, A. von Gottberg, R. S. Heyderman, N. French, A. Kadioglu, W. P. Hanage, D. B. Everett, S. D. Bentley, and P. AGE Consortium. 2016. 'Understanding pneumococcal serotype 1 biology through population genomic analysis', *BMC Infect Dis*, 16: 649.
- Chaguza, C., C. Ebruke, M. Senghore, S. W. Lo, P. E. Tientcheu, R. A. Gladstone, G. Tonkin-Hill, J. E. Cornick, M. Yang, A. Worwui, L. McGee, R. F. Breiman, K. P. Klugman, A. Kadioglu, D. B. Everett, G. Mackenzie, N. J. Croucher, A. Roca, B. A. Kwambana-Adams, M. Antonio, and S. D. Bentley. 2022. 'Comparative genomics of disease and carriage serotype 1 pneumococci', *Genome Biol Evol*.
- Chang, J. C., B. LaSarre, J. C. Jimenez, C. Aggarwal, and M. J. Federle. 2011. 'Two group A streptococcal peptide pheromones act through opposing Rgg regulators to control biofilm development', *PLoS Pathog*, 7: e1002190.
- Chang, X., Y. Li, J. Ping, X. B. Xing, H. Sun, P. Jia, C. Wang, Y. Y. Li, and Y. X. Li. 2011. 'EcoBrowser: a web-based tool for visualizing transcriptome data of *Escherichia coli*', *BMC Res Notes*, 4: 405.

- Chao, Y., C. Bergenfelz, R. Sun, X. Han, A. Achour, and A. P. Hakansson. 2020. 'The serine protease HtrA plays a key role in heat-induced dispersal of pneumococcal biofilms', *Sci Rep*, 10: 22455.
- Charrad, Malika, Nadia Ghazzali, Véronique Boiteau, and Azam Niknafs. 2014. 'NbClust: AnRPackage for Determining the Relevant Number of Clusters in a Data Set', *Journal of Statistical Software*, 61.
- Chiavolini, D., G. Pozzi, and S. Ricci. 2008. 'Animal models of Streptococcus pneumoniae disease', *Clin Microbiol Rev*, 21: 666-85.
- Chin, C. S., D. H. Alexander, P. Marks, A. A. Klammer, J. Drake, C. Heiner, A. Clum, A. Copeland, J. Huddleston, E. E. Eichler, S. W. Turner, and J. Korlach. 2013. 'Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data', *Nat Methods*, 10: 563-9.
- Chou, W. C., Q. Ma, S. Yang, S. Cao, D. M. Klingeman, S. D. Brown, and Y. Xu. 2015. 'Analysis of strand-specific RNA-seq data using machine learning reveals the structures of transcription units in Clostridium thermocellum', *Nucleic Acids Res*, 43: e67.
- Clark, S. A., R. Doyle, J. Lucidarme, R. Borrow, and J. Breuer. 2018. 'Targeted DNA enrichment and whole genome sequencing of Neisseria meningitidis directly from clinical specimens', *Int J Med Microbiol*, 308: 256-62.
- Claverys, J. P., M. Prudhomme, and B. Martin. 2006. 'Induction of competence regulons as a general response to stress in gram-positive bacteria', *Annu Rev Microbiol*, 60: 451-75.
- Colgan, A. M., A. D. Cameron, and C. Kroger. 2017. 'If it transcribes, we can sequence it: mining the complexities of host-pathogen-environment interactions using RNA-seq', *Curr Opin Microbiol*, 36: 37-46.
- Corbin, B. D., E. H. Seeley, A. Raab, J. Feldmann, M. R. Miller, V. J. Torres, K. L. Anderson, B. M. Dattilo, P. M. Dunman, R. Gerads, R. M. Caprioli, W. Nacken, W. J. Chazin, and E. P.

- Skaar. 2008. 'Metal chelation and inhibition of bacterial growth in tissue abscesses', *Science*, 319: 962-5.
- Cross, J. H., R. S. Bradbury, A. J. Fulford, A. T. Jallow, R. Wegmuller, A. M. Prentice, and C. Cerami. 2015. 'Oral iron acutely elevates bacterial growth in human serum', *Sci Rep*, 5: 16670.
- Croucher, N. J., and N. R. Thomson. 2010. 'Studying bacterial transcriptomes using RNA-seq', *Curr Opin Microbiol*, 13: 619-24.
- Crowley, P. J., J. A. Gutierrez, J. D. Hillman, and A. S. Bleiweis. 1997. 'Genetic and physiologic analysis of a formyl-tetrahydrofolate synthetase mutant of *Streptococcus mutans*', *J Bacteriol*, 179: 1563-72.
- Cunningham-Rundles, S., P. J. Giardina, R. W. Grady, C. Califano, P. McKenzie, and M. De Sousa. 2000. 'Effect of transfusional iron overload on immune response', *J Infect Dis*, 182 Suppl 1: S115-21.
- D'Mello, A., A. N. Riegler, E. Martinez, S. M. Beno, T. D. Ricketts, E. F. Foxman, C. J. Orihuela, and H. Tettelin. 2020. 'An in vivo atlas of host-pathogen transcriptomes during *Streptococcus pneumoniae* colonization and disease', *Proc Natl Acad Sci U S A*, 117: 33507-18.
- De Coster, W., S. D'Hert, D. T. Schultz, M. Cruts, and C. Van Broeckhoven. 2018. 'NanoPack: visualizing and processing long-read sequencing data', *Bioinformatics*, 34: 2666-69.
- de Stoppelaar, S. F., H. J. Bootsma, A. Zomer, J. J. Roelofs, P. W. Hermans, C. van 't Veer, and T. van der Poll. 2013. '*Streptococcus pneumoniae* serine protease HtrA, but not SFP or PrtA, is a major virulence factor in pneumonia', *PLoS One*, 8: e80062.
- Delekta, P. C., J. C. Shook, T. A. Lydic, M. H. Mulks, and N. D. Hammer. 2018. '*Staphylococcus aureus* Utilizes Host-Derived Lipoprotein Particles as Sources of Fatty Acids', *J Bacteriol*, 200.

- Denoel, P., F. Godfroid, P. Hermand, V. Verlant, and J. Poolman. 2011. 'Combined protective effects of anti-PhtD and anti-pneumococcal polysaccharides', *Vaccine*, 29: 6451-3.
- Djoko, K. Y., C. L. Ong, M. J. Walker, and A. G. McEwan. 2015. 'The Role of Copper and Zinc Toxicity in Innate Immune Defense against Bacterial Pathogens', *J Biol Chem*, 290: 18954-61.
- Donati, C., N. L. Hiller, H. Tettelin, A. Muzzi, N. J. Croucher, S. V. Angiuoli, M. Oggioni, J. C. Dunning Hotopp, F. Z. Hu, D. R. Riley, A. Covacci, T. J. Mitchell, S. D. Bentley, M. Kilian, G. D. Ehrlich, R. Rappuoli, E. R. Moxon, and V. Masignani. 2010. 'Structure and dynamics of the pan-genome of *Streptococcus pneumoniae* and closely related species', *Genome Biol*, 11: R107.
- Duong, N., S. Osborne, V. H. Bustamante, A. M. Tomljenovic, J. L. Puente, and B. K. Coombes. 2007. 'Thermosensing coordinates a cis-regulatory module for transcriptional activation of the intracellular virulence system in *Salmonella enterica* serovar Typhimurium', *J Biol Chem*, 282: 34077-84.
- Eastel, J. M., K. W. Lam, N. L. Lee, W. Y. Lok, A. H. F. Tsang, X. M. Pei, A. K. C. Chan, W. C. S. Cho, and S. C. C. Wong. 2019. 'Application of NanoString technologies in companion diagnostic development', *Expert Rev Mol Diagn*, 19: 591-98.
- Echave, P., J. Tamarit, E. Cabiscol, and J. Ros. 2003. 'Novel antioxidant role of alcohol dehydrogenase E from *Escherichia coli*', *J Biol Chem*, 278: 30193-8.
- Eichner, H., J. Karlsson, L. Spelmink, A. Pathak, L. T. Sham, B. Henriques-Normark, and E. Loh. 2021. 'RNA thermosensors facilitate *Streptococcus pneumoniae* and *Haemophilus influenzae* immune evasion', *PLoS Pathog*, 17: e1009513.
- Eid, J., A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto, P. Peluso, D. Rank, P. Baybayan, B. Bettman, A. Bibillo, K. Bjornson, B. Chaudhuri, F. Christians, R. Cicero, S. Clark, R. Dalal, A. Dewinter, J. Dixon, M. Foquet, A. Gaertner, P. Hardenbol, C. Heiner, K. Hester, D. Holden,

G. Kearns, X. Kong, R. Kuse, Y. Lacroix, S. Lin, P. Lundquist, C. Ma, P. Marks, M. Maxham, D. Murphy, I. Park, T. Pham, M. Phillips, J. Roy, R. Sebra, G. Shen, J. Sorenson, A. Tomaney, K. Travers, M. Trulson, J. Vieceli, J. Wegener, D. Wu, A. Yang, D. Zaccarin, P. Zhao, F. Zhong, J. Korlach, and S. Turner. 2009. 'Real-time DNA sequencing from single polymerase molecules', *Science*, 323: 133-8.

Eisen, M. B., P. T. Spellman, P. O. Brown, and D. Botstein. 1998. 'Cluster analysis and display of genome-wide expression patterns', *Proc Natl Acad Sci U S A*, 95: 14863-8.

Ewels, P., M. Magnusson, S. Lundin, and M. Kaller. 2016. 'MultiQC: summarize analysis results for multiple tools and samples in a single report', *Bioinformatics*, 32: 3047-8.

Feldman, C., and R. Anderson. 2020. 'Recent advances in the epidemiology and prevention of Streptococcus pneumoniae infections', *F1000Res*, 9.

Ferreira, D. M., D. R. Neill, M. Bangert, J. F. Gritzfeld, N. Green, A. K. Wright, S. H. Pennington, L. Bricio-Moreno, A. T. Moreno, E. N. Miyaji, A. D. Wright, A. M. Collins, D. Goldblatt, A. Kadioglu, and S. B. Gordon. 2013. 'Controlled human infection and rechallenge with Streptococcus pneumoniae reveals the protective efficacy of carriage in healthy adults', *Am J Respir Crit Care Med*, 187: 855-64.

Filannino, P., M. De Angelis, R. Di Cagno, G. Gozzi, Y. Riciputi, and M. Gobbetti. 2018. 'How Lactobacillus plantarum shapes its transcriptome in response to contrasting habitats', *Environ Microbiol*, 20: 3700-16.

Fraenkel, Albert. 1886. *Bakteriologische mittheilungen*.

Gamez, G., A. Castro, A. Gomez-Mejia, M. Gallego, A. Bedoya, M. Camargo, and S. Hammerschmidt. 2018. 'The variome of pneumococcal virulence factors and regulators', *BMC Genomics*, 19: 10.

Ganaie, F., J. S. Saad, L. McGee, A. J. van Tonder, S. D. Bentley, S. W. Lo, R. A. Gladstone, P. Turner, J. D. Keenan, R. F. Breiman, and M. H. Nahm. 2020. 'A New Pneumococcal



Capsule Type, 10D, is the 100th Serotype and Has a Large cps Fragment from an Oral Streptococcus', *mBio*, 11.

Garzetti, D., S. Brugiroux, B. Bunk, R. Pukall, K. D. McCoy, A. J. Macpherson, and B. Stecher. 2017. 'High-Quality Whole-Genome Sequences of the Oligo-Mouse-Microbiota Bacterial Community', *Genome Announc*, 5.

Geiss, G. K., R. E. Bumgarner, B. Birditt, T. Dahl, N. Dowidar, D. L. Dunaway, H. P. Fell, S. Ferree, R. D. George, T. Grogan, J. J. James, M. Maysuria, J. D. Mitton, P. Oliveri, J. L. Osborn, T. Peng, A. L. Ratcliffe, P. J. Webster, E. H. Davidson, L. Hood, and K. Dimitrov. 2008. 'Direct multiplexed measurement of gene expression with color-coded probe pairs', *Nat Biotechnol*, 26: 317-25.

Geissler, A. S., C. Anthon, F. Alkan, E. Gonzalez-Tortuero, L. D. Poulsen, T. B. Kallehauge, A. Breuner, S. E. Seemann, J. Vinther, and J. Gorodkin. 2021. 'BSGatlas: a unified *Bacillus subtilis* genome and transcriptome annotation atlas with enhanced information access', *Microb Genom*, 7.

Glanville, D. G., L. Han, A. F. Maule, A. Woodacre, D. Thanki, I. T. Abdullah, J. A. Morrissey, T. B. Clarke, H. Yesilkaya, N. R. Silvaggi, and A. T. Ulijasz. 2018. 'RitR is an archetype for a novel family of redox sensors in the streptococci that has evolved from two-component response regulators and is required for pneumococcal colonization', *PLoS Pathog*, 14: e1007052.

Goeman, J. J., and P. Buhlmann. 2007. 'Analyzing gene expression data in terms of gene sets: methodological issues', *Bioinformatics*, 23: 980-7.

Gopal, T., V. Nagarajan, and M. O. Elasri. 2015. 'SATRAT: *Staphylococcus aureus* transcript regulatory network analysis tool', *PeerJ*, 3: e717.

Granat, S. M., J. Ollgren, E. Herva, Z. Mia, K. Auranen, and P. H. Makela. 2009. 'Epidemiological evidence for serotype-independent acquired immunity to pneumococcal carriage', *J Infect Dis*, 200: 99-106.

Greenwood, Brian. 1999. 'Meningococcal meningitis in Africa', *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 93: 341-53.

Griffith, F. 1928. 'The Significance of Pneumococcal Types', *J Hyg (Lond)*, 27: 113-59.

Gritzfeld, J. F., A. D. Wright, A. M. Collins, S. H. Pennington, A. K. Wright, A. Kadioglu, D. M. Ferreira, and S. B. Gordon. 2013. 'Experimental human pneumococcal carriage', *J Vis Exp*.

Guiral, S., V. Henard, C. Granadel, B. Martin, and J. P. Claverys. 2006. 'Inhibition of competence development in *Streptococcus pneumoniae* by increased basal-level expression of the ComDE two-component regulatory system', *Microbiology (Reading)*, 152: 323-31.

Gurevich, A., V. Saveliev, N. Vyahhi, and G. Tesler. 2013. 'QUAST: quality assessment tool for genome assemblies', *Bioinformatics*, 29: 1072-5.

Haas, B. J., M. Chin, C. Nusbaum, B. W. Birren, and J. Livny. 2012. 'How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes?', *BMC Genomics*, 13: 734.

Hakansson, A. P. 2014. 'Pneumococcal adaptive responses to changing host environments', *J Infect Dis*, 210: 1-3.

Hammarlof, D. L., C. Kroger, S. V. Owen, R. Canals, L. Lacharme-Lora, N. Wenner, A. E. Schager, T. J. Wells, I. R. Henderson, P. Wigley, K. Hokamp, N. A. Feasey, M. A. Gordon, and J. C. D. Hinton. 2018. 'Role of a single noncoding nucleotide in the evolution of an epidemic African clade of *Salmonella*', *Proc Natl Acad Sci U S A*, 115: E2614-E23.

Hammit, L. L., J. C. Campbell, D. Borys, R. C. Weatherholtz, R. Reid, N. Goklish, L. H. Moulton, M. Traskine, Y. Song, K. Swinnen, M. Santosham, and K. L. O'Brien. 2019. 'Efficacy, safety and immunogenicity of a pneumococcal protein-based vaccine co-administered with 13-valent pneumococcal conjugate vaccine against acute otitis media in young children: A phase IIb randomized study', *Vaccine*, 37: 7482-92.

- Hathaway, L. J., S. D. Brugger, B. Morand, M. Bangert, J. U. Rotzetter, C. Hauser, W. A. Graber, S. Gore, A. Kadioglu, and K. Muhlemann. 2012. 'Capsule type of *Streptococcus pneumoniae* determines growth phenotype', *PLoS Pathog*, 8: e1002574.
- Hava, D. L., J. LeMieux, and A. Camilli. 2003. 'From nose to lung: the regulation behind *Streptococcus pneumoniae* virulence factors', *Mol Microbiol*, 50: 1103-10.
- Hava, David L., and Andrew Camilli. 2002. 'Large-scale identification of serotype 4 *Streptococcus pneumoniae* virulence factors', *Molecular Microbiology*, 45: 1389-406.
- Havarstein, L. S., G. Coomaraswamy, and D. A. Morrison. 1995. 'An unmodified heptadecapeptide pheromone induces competence for genetic transformation in *Streptococcus pneumoniae*', *Proc Natl Acad Sci U S A*, 92: 11140-4.
- Heffron, Roderick. 1939. 'Pneumonia, with General Reference to *Pneumococcus Lobar Pneumonia*', *Oxford 1939*.
- Henriques-Normark, B., and E. I. Tuomanen. 2013. 'The pneumococcus: epidemiology, microbiology, and pathogenesis', *Cold Spring Harb Perspect Med*, 3.
- Hill, P. C., A. Akisanya, K. Sankareh, Y. B. Cheung, M. Saaka, G. Lahai, B. M. Greenwood, and R. A. Adegbola. 2006. 'Nasopharyngeal carriage of *Streptococcus pneumoniae* in Gambian villagers', *Clin Infect Dis*, 43: 673-9.
- Hiller, N. L., B. Janto, J. S. Hogg, R. Boissy, S. Yu, E. Powell, R. Keefe, N. E. Ehrlich, K. Shen, J. Hayes, K. Barbadora, W. Klimke, D. Dernovoy, T. Tatusova, J. Parkhill, S. D. Bentley, J. C. Post, G. D. Ehrlich, and F. Z. Hu. 2007. 'Comparative genomic analyses of seventeen *Streptococcus pneumoniae* strains: insights into the pneumococcal supragenome', *J Bacteriol*, 189: 8186-95.
- Hoe, N. P., and J. D. Goguen. 1993. 'Temperature sensing in *Yersinia pestis*: translation of the LcrF activator protein is thermally regulated', *J Bacteriol*, 175: 7901-9.

Holyer, R. H., C. D. Hubbard, S. F. A. Kettle, and R. G. Wilkins. 2002. 'The Kinetics of Replacement Reactions of Complexes of the Transition Metals with 1,10-Phenanthroline and 2,2'-Bipyridine', *Inorganic Chemistry*, 4: 929-35.

Honsa, E. S., M. D. Johnson, and J. W. Rosch. 2013. 'The roles of transition metals in the physiology and pathogenesis of *Streptococcus pneumoniae*', *Front Cell Infect Microbiol*, 3: 92.

Hood, M. I., and E. P. Skaar. 2012. 'Nutritional immunity: transition metals at the pathogen-host interface', *Nat Rev Microbiol*, 10: 525-37.

Hoskins, J., W. E. Alborn, Jr., J. Arnold, L. C. Blaszcak, S. Burgett, B. S. DeHoff, S. T. Estrem, L. Fritz, D. J. Fu, W. Fuller, C. Geringer, R. Gilmour, J. S. Glass, H. Khoja, A. R. Kraft, R. E. Lagace, D. J. LeBlanc, L. N. Lee, E. J. Lefkowitz, J. Lu, P. Matsushima, S. M. McAhren, M. McHenney, K. McLeaster, C. W. Mundy, T. I. Nicas, F. H. Norris, M. O'Gara, R. B. Peery, G. T. Robertson, P. Rockey, P. M. Sun, M. E. Winkler, Y. Yang, M. Young-Bellido, G. Zhao, C. A. Zook, R. H. Baltz, S. R. Jaskunas, P. R. Rosteck, Jr., P. L. Skatrud, and J. I. Glass. 2001. 'Genome of the bacterium *Streptococcus pneumoniae* strain R6', *J Bacteriol*, 183: 5709-17.

Hotomi, M., J. Yuasa, D. E. Briles, and N. Yamanaka. 2016. 'Pneumolysin plays a key role at the initial step of establishing pneumococcal nasal colonization', *Folia Microbiol (Praha)*, 61: 375-83.

Hoyer, J., J. Bartel, A. Gomez-Mejia, M. Rohde, C. Hirschfeld, N. Hess, T. Sura, S. Maass, S. Hammerschmidt, and D. Becher. 2018. 'Proteomic response of *Streptococcus pneumoniae* to iron limitation', *Int J Med Microbiol*, 308: 713-21.

Hua, C. Z., A. Howard, R. Malley, and Y. J. Lu. 2014. 'Effect of nonheme iron-containing ferritin Dpr in the stress response and virulence of pneumococci', *Infect Immun*, 82: 3939-47.

- Huang, S. T., C. L. Lin, Y. J. Chang, Y. P. Sher, M. J. Wu, K. H. Shu, F. C. Sung, and C. H. Kao. 2014. 'Pneumococcal pneumonia infection is associated with end-stage renal disease in adult hospitalized patients', *Kidney Int*, 86: 1023-30.
- Hurme, Reini, Kurt D. Berndt, Staffan J. Normark, and Mikael Rhen. 1997. 'A Proteinaceous Gene Regulatory Thermometer in Salmonella', *Cell*, 90: 55-64.
- Hyams, C., E. Camberlein, J. M. Cohen, K. Bax, and J. S. Brown. 2010. 'The Streptococcus pneumoniae capsule inhibits complement activity and neutrophil phagocytosis by multiple mechanisms', *Infect Immun*, 78: 704-15.
- Iannelli, F., M. R. Oggioni, and G. Pozzi. 2005. 'Sensor domain of histidine kinase ComD confers competence pherotype specificity in Streptococcus pneumoniae', *FEMS Microbiol Lett*, 252: 321-6.
- Iannelli, Francesco, Marco R. Oggioni, and Gianni Pozzi. 2002. 'Allelic variation in the highly polymorphic locus pspC of Streptococcus pneumoniae', *Gene*, 284: 63-71.
- Ibrahim, Y. M., A. R. Kerr, J. McCluskey, and T. J. Mitchell. 2004. 'Role of HtrA in the virulence and competence of Streptococcus pneumoniae', *Infect Immun*, 72: 3584-91.
- Igartua, C., E. R. Davenport, Y. Gilad, D. L. Nicolae, J. Pinto, and C. Ober. 2017. 'Host genetic variation in mucosal immunity pathways influences the upper airway microbiome', *Microbiome*, 5: 16.
- Iyer, R., and A. Camilli. 2007. 'Sucrose metabolism contributes to in vivo fitness of Streptococcus pneumoniae', *Mol Microbiol*, 66: 1-13.
- Jacques, L. C., S. Panagiotou, M. Baltazar, M. Senghore, S. Khandaker, R. Xu, L. Bricio-Moreno, M. Yang, C. G. Dowson, D. B. Everett, D. R. Neill, and A. Kadioglu. 2020. 'Increased pathogenicity of pneumococcal serotype 1 is driven by rapid autolysis and release of pneumolysin', *Nat Commun*, 11: 1892.

Jamalkandi, S. A., M. Kouhsar, J. Salimian, and A. Ahmadi. 2020. 'The identification of co-expressed gene modules in *Streptococcus pneumonia* from colonization to infection to predict novel potential virulence genes', *BMC Microbiol*, 20: 376.

Jedrzejewski, M. J. 2001. 'Pneumococcal virulence factors: structure and function', *Microbiol Mol Biol Rev*, 65: 187-207 ; first page, table of contents.

Jerga, A., and C. O. Rock. 2009. 'Acyl-Acyl carrier protein regulates transcription of fatty acid biosynthetic genes via the FabT repressor in *Streptococcus pneumoniae*', *J Biol Chem*, 284: 15364-8.

Jimenez-Munguia, I., M. Calderon-Santiago, A. Rodriguez-Franco, F. Priego-Capote, and M. J. Rodriguez-Ortega. 2018. 'Multi-omic profiling to assess the effect of iron starvation in *Streptococcus pneumoniae* TIGR4', *PeerJ*, 6: e4966.

Jochems, S. P., K. de Ruiter, C. Solorzano, A. Voskamp, E. Mitsi, E. Nikolaou, B. F. Carniel, S. Pojar, E. L. German, J. Reine, A. Soares-Schanoski, H. Hill, R. Robinson, A. D. Hyder-Wright, C. M. Weight, P. F. Durrenberger, R. S. Heyderman, S. B. Gordon, H. H. Smits, B. C. Urban, J. Rylance, A. M. Collins, M. D. Wilkie, L. Lazarova, S. C. Leong, M. Yazdanbakhsh, and D. M. Ferreira. 2019. 'Innate and adaptive nasal mucosal immune responses following experimental human pneumococcal colonization', *J Clin Invest*, 129: 4523-38.

Johansson, Jörgen, Pierre Mandin, Adriana Renzoni, Claude Chiaruttini, Mathias Springer, and Pascale Cossart. 2002. 'An RNA Thermosensor Controls Expression of Virulence Genes in *Listeria monocytogenes*', *Cell*, 110: 551-61.

Johnston, J. W., D. E. Briles, L. E. Myers, and S. K. Hollingshead. 2006. 'Mn<sup>2+</sup>-dependent regulation of multiple genes in *Streptococcus pneumoniae* through PsaR and the resultant impact on virulence', *Infect Immun*, 74: 1171-80.

Jolley, K. A., J. E. Bray, and M. C. J. Maiden. 2018. 'Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications', *Wellcome Open Res*, 3: 124.

- Jomaa, M., S. Terry, C. Hale, C. Jones, G. Dougan, and J. Brown. 2006. 'Immunization with the iron uptake ABC transporter proteins PiaA and PiuA prevents respiratory infection with *Streptococcus pneumoniae*', *Vaccine*, 24: 5133-9.
- Jones, C. H., T. C. Bolken, K. F. Jones, G. O. Zeller, and D. E. Hruby. 2001. 'Conserved DegP protease in gram-positive bacteria is essential for thermal and oxidative tolerance and full virulence in *Streptococcus pyogenes*', *Infect Immun*, 69: 5538-45.
- Jusot, J. F., D. R. Neill, E. M. Waters, M. Bangert, M. Collins, L. Bricio Moreno, K. G. Lawan, M. M. Moussa, E. Dearing, D. B. Everett, J. M. Collard, and A. Kadioglu. 2017. 'Airborne dust and high temperatures are risk factors for invasive bacterial disease', *J Allergy Clin Immunol*, 139: 977-86 e2.
- Kalizang'oma, A., C. Chaguza, A. Gori, C. Davison, S. Beleza, M. Antonio, B. Beall, D. Goldblatt, B. Kwambana-Adams, S. D. Bentley, and R. S. Heyderman. 2021. 'Streptococcus pneumoniae serotypes that frequently colonise the human nasopharynx are common recipients of penicillin-binding protein gene fragments from *Streptococcus mitis*', *Microb Genom*, 7.
- Kalvari, I., E. P. Nawrocki, J. Argasinska, N. Quinones-Olvera, R. D. Finn, A. Bateman, and A. I. Petrov. 2018. 'Non-Coding RNA Analysis Using the Rfam Database', *Curr Protoc Bioinformatics*, 62: e51.
- Kamada, M., S. Hase, K. Sato, A. Toyoda, A. Fujiyama, and Y. Sakakibara. 2014. 'Whole genome complete resequencing of *Bacillus subtilis* natto by combining long reads with high-quality short reads', *PLoS One*, 9: e109999.
- Kanehisa, M. 2000. 'KEGG: Kyoto Encyclopedia of Genes and Genomes', *Nucleic Acids Research*, 28: 27-30.
- Kapatai, G., C. L. Sheppard, A. Al-Shahib, D. J. Litt, A. P. Underwood, T. G. Harrison, and N. K. Fry. 2016. 'Whole genome sequencing of *Streptococcus pneumoniae*: development,

evaluation and verification of targets for serogroup and serotype prediction using an automated pipeline', *PeerJ*, 4: e2477.

Kaur, G., and J. M. Dufour. 2012. 'Cell lines: Valuable tools or useless artifacts', *Spermatogenesis*, 2: 1-5.

Kehl-Fie, T. E., and E. P. Skaar. 2010. 'Nutritional immunity beyond iron: a role for manganese and zinc', *Curr Opin Chem Biol*, 14: 218-24.

Keller, L. E., D. A. Robinson, and L. S. McDaniel. 2016. 'Nonencapsulated *Streptococcus pneumoniae*: Emergence and Pathogenesis', *mBio*, 7: e01792.

King, S. J. 2010. 'Pneumococcal modification of host sugars: a major contributor to colonization of the human airway?', *Mol Oral Microbiol*, 25: 15-24.

Kjos, M., E. Miller, J. Slager, F. B. Lake, O. Gericke, I. S. Roberts, D. E. Rozen, and J. W. Veening. 2016. 'Expression of *Streptococcus pneumoniae* Bacteriocins Is Induced by Antibiotics via Regulatory Interplay with the Competence System', *PLoS Pathog*, 12: e1005422.

Klemperer, Georg, and Felix Klemperer. 1891a. *Versuche über Immunisierung und Heilung bei der Pneumokokkeninfection* (Berliner Klinische Wochenschrift).

Klinkert, B., A. Cimdins, L. C. Gaubig, J. Rossmann, U. Aschke-Sonnenborn, and F. Narberhaus. 2012. 'Thermogenetic tools to monitor temperature-dependent gene expression in bacteria', *J Biotechnol*, 160: 55-63.

Klinkert, B., and F. Narberhaus. 2009. 'Microbial thermosensors', *Cell Mol Life Sci*, 66: 2661-76.

Koren, S., G. P. Harhay, T. P. Smith, J. L. Bono, D. M. Harhay, S. D. McVey, D. Radune, N. H. Bergman, and A. M. Phillippy. 2013. 'Reducing assembly complexity of microbial genomes with single-molecule sequencing', *Genome Biol*, 14: R101.



- Kroger, C., A. Colgan, S. Srikumar, K. Handler, S. K. Sivasankaran, D. L. Hammarlof, R. Canals, J. E. Grissom, T. Conway, K. Hokamp, and J. C. Hinton. 2013. 'An infection-relevant transcriptomic compendium for *Salmonella enterica* Serovar Typhimurium', *Cell Host Microbe*, 14: 683-95.
- Kroger, C., K. D. MacKenzie, E. Y. Alshabib, M. W. B. Kirzinger, D. M. Suchan, T. C. Chao, V. Akulova, A. A. Miranda-CasoLuengo, V. A. Monzon, T. Conway, S. K. Sivasankaran, J. C. D. Hinton, K. Hokamp, and A. D. S. Cameron. 2018. 'The primary transcriptome, small RNAs and regulation of antimicrobial resistance in *Acinetobacter baumannii* ATCC 17978', *Nucleic Acids Res*, 46: 9684-98.
- Kulohoma, B. W., J. E. Cornick, C. Chaguza, F. Yalcin, S. R. Harris, K. J. Gray, A. M. Kiran, E. Molyneux, N. French, J. Parkhill, B. E. Faragher, D. B. Everett, S. D. Bentley, and R. S. Heyderman. 2015. 'Comparative Genomic Analysis of Meningitis- and Bacteremia-Causing *Pneumococci* Identifies a Common Core Genome', *Infect Immun*, 83: 4165-73.
- Kwambana-Adams, B. A., F. Asiedu-Bekoe, B. Sarkodie, O. K. Afreh, G. K. Kuma, G. Owusu-Okyere, E. Foster-Nyarko, S. A. Ohene, C. Okot, A. K. Worwui, C. Okoi, M. Senghore, J. K. Otu, C. Ebruke, R. Bannerman, K. Amponsa-Achiano, D. Opare, G. Kay, T. Letsa, O. Kaluwa, E. Appiah-Denkyira, V. Bampoe, S. M. Zaman, M. J. Pallen, U. D'Alessandro, J. M. Mwenda, and M. Antonio. 2016. 'An outbreak of pneumococcal meningitis among older children ( $\geq 5$  years) and adults after the implementation of an infant vaccination programme with the 13-valent pneumococcal conjugate vaccine in Ghana', *BMC Infect Dis*, 16: 575.
- Kwon, H. Y., S. W. Kim, M. H. Choi, A. D. Ogunniyi, J. C. Paton, S. H. Park, S. N. Pyo, and D. K. Rhee. 2003. 'Effect of heat shock and mutations in ClpL and ClpP on virulence gene expression in *Streptococcus pneumoniae*', *Infect Immun*, 71: 3757-65.
- Lamble, S., E. Batty, M. Attar, D. Buck, R. Bowden, G. Lunter, D. Crook, B. El-Fahmawi, and P. Piazza. 2013. 'Improved workflows for high throughput library preparation using the transposome-based Nextera system', *BMC Biotechnol*, 13: 104.

- Langmead, B., and S. L. Salzberg. 2012. 'Fast gapped-read alignment with Bowtie 2', *Nat Methods*, 9: 357-9.
- Lanie, J. A., W. L. Ng, K. M. Kazmierczak, T. M. Andrzejewski, T. M. Davidsen, K. J. Wayne, H. Tettelin, J. I. Glass, and M. E. Winkler. 2007. 'Genome sequence of Avery's virulent serotype 2 strain D39 of *Streptococcus pneumoniae* and comparison with that of unencapsulated laboratory strain R6', *J Bacteriol*, 189: 38-51.
- Lau, G. W., S. Haataja, M. Lonetto, S. E. Kensit, A. Marra, A. P. Bryant, D. McDevitt, D. A. Morrison, and D. W. Holden. 2001. 'A functional genomic analysis of type 3 *Streptococcus pneumoniae* virulence', *Mol Microbiol*, 40: 555-71.
- Lees, J. A., N. J. Croucher, D. Goldblatt, F. Nosten, J. Parkhill, C. Turner, P. Turner, and S. D. Bentley. 2017. 'Genome-wide identification of lineage and locus specific variation associated with pneumococcal carriage duration', *Elife*, 6.
- Leimkugel, J., A. Adams Forgor, S. Gagneux, V. Pfluger, C. Flierl, E. Awine, M. Naegeli, J. P. Dangy, T. Smith, A. Hodgson, and G. Pluschke. 2005. 'An outbreak of serotype 1 *Streptococcus pneumoniae* meningitis in northern Ghana with features that are characteristic of *Neisseria meningitidis* meningitis epidemics', *J Infect Dis*, 192: 192-9.
- Lessa, F. C., J. Milucky, N. G. Rouphael, N. M. Bennett, H. K. Talbot, L. H. Harrison, M. M. Farley, J. Walston, F. Pimenta, R. E. Gertz, G. Rajam, M. D. G. Carvalho, B. Beall, and C. G. Whitney. 2018. 'Streptococcus mitis Expressing Pneumococcal Serotype 1 Capsule', *Sci Rep*, 8: 17959.
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and Subgroup Genome Project Data Processing. 2009. 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics*, 25: 2078-9.
- Li, J. W., J. Li, J. Wang, C. Li, and J. R. Zhang. 2019. 'Molecular Mechanisms of hsdS Inversions in the cod Locus of *Streptococcus pneumoniae*', *J Bacteriol*, 201.

- Liao, Y., G. K. Smyth, and W. Shi. 2014. 'featureCounts: an efficient general purpose program for assigning sequence reads to genomic features', *Bioinformatics*, 30: 923-30.
- Lindemann, J., R. Leiacker, G. Rettinger, and T. Keck. 2002. 'Nasal mucosal temperature during respiration', *Clinical Otolaryngology and Allied Sciences*, 27: 135-39.
- Lipinska, B., O. Fayet, L. Baird, and C. Georgopoulos. 1989. 'Identification, characterization, and mapping of the Escherichia coli htrA gene, whose product is essential for bacterial growth only at elevated temperatures', *J Bacteriol*, 171: 1574-84.
- Liu, X., J. M. Kimmey, L. Matarazzo, V. de Bakker, L. Van Maele, J. C. Sirard, V. Nizet, and J. W. Veening. 2021. 'Exploration of Bacterial Bottlenecks and Streptococcus pneumoniae Pathogenesis by CRISPRi-Seq', *Cell Host Microbe*, 29: 107-20 e6.
- Liu, Y., A. Beyer, and R. Aebersold. 2016. 'On the Dependency of Cellular Protein Levels on mRNA Abundance', *Cell*, 165: 535-50.
- Lo, Stephanie W., Rebecca A. Gladstone, Andries J. van Tonder, John A. Lees, Mignon du Plessis, Rachel Benisty, Noga Givon-Lavi, Paulina A. Hawkins, Jennifer E. Cornick, Brenda Kwambana-Adams, Pierra Y. Law, Pak Leung Ho, Martin Antonio, Dean B. Everett, Ron Dagan, Anne von Gottberg, Keith P. Klugman, Lesley McGee, Robert F. Breiman, Stephen D. Bentley, Abdullah W. Brooks, Alejandra Corso, Alexander Davydov, Alison Maguire, Andrew Pollard, Anmol Kiran, Anna Skoczynska, Benild Moiane, Bernard Beall, Betuel Sigauque, David Aanensen, Deborah Lehmann, Diego Faccone, Ebenezer Foster-Nyarko, Ebrima Bojang, Ekaterina Egorova, Elena Voropaeva, Eric Sampane-Donkor, Ewa Sadowy, Godfrey Bigogo, Helio Mucavele, Houria Belabbès, Idrissa Diawara, Jennifer Moïsi, Jennifer Verani, Jeremy Keenan, Jyothish N. Nair Thulasee Bhai, Kedibone M. Ndlangisa, Khalid Zerouali, K. L. Ravikumar, Leonid Titov, Linda De Gouveia, Maaïke Alaerts, Margaret Ip, Maria Cristina de Cunto Brandileone, Md Hasanuzzaman, Metka Paragi, Michele Nurse-Lucas, Mushal Ali, Naima Elmdaghri, Nicholas Croucher, Nicole Wolter, Nurit Porat, Özgen Köseoglu Eser, Patrick E. Akpaka, Paul Turner, Paula Galletti, Peggy-Estelle Tientcheu, Philip E. Carter, Rafal Mostowy, Rama Kandasamy, Rebecca Ford, Rebecca Henderson,

- Roly Malaker, Sadia Shakoar, Samanta Cristine Grassi Almeida, Samir K. Saha, Sanjay Doiphode, Shabir A. Madhi, Shamala Devi Sekaran, Somporn Srifuengfung, Stephen Obaro, Stuart C. Clarke, Susan A. Nzenze, Tamara Kastrin, Theresa J. Ochoa, Veeraraghavan Balaji, Waleria Hryniewicz, and Yulia Urban. 2019. 'Pneumococcal lineages associated with serotype replacement and antibiotic resistance in childhood invasive pneumococcal disease in the post-PCV13 era: an international whole-genome sequencing study', *The Lancet Infectious Diseases*, 19: 759-69.
- Loh, E., E. Kugelberg, A. Tracy, Q. Zhang, B. Gollan, H. Ewles, R. Chalmers, V. Pelicic, and C. M. Tang. 2013. 'Temperature triggers immune evasion by *Neisseria meningitidis*', *Nature*, 502: 237-40.
- Loh, E., F. Righetti, H. Eichner, C. Twittenhoff, and F. Narberhaus. 2018. 'RNA Thermometers in Bacterial Pathogens', *Microbiol Spectr*, 6.
- Lopez-Agudelo, V. A., A. Baena, V. Barrera, F. Cabarcas, J. F. Alzate, D. J. V. Beste, R. Rios-Estapa, and L. F. Barrera. 2022. 'Dual RNA Sequencing of *Mycobacterium tuberculosis*-Infected Human Splenic Macrophages Reveals a Strain-Dependent Host-Pathogen Response to Infection', *Int J Mol Sci*, 23.
- Love, M. I., W. Huber, and S. Anders. 2014. 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biol*, 15: 550.
- Mahdi, L. K., H. Wang, M. B. Van der Hoek, J. C. Paton, and A. D. Ogunniyi. 2012. 'Identification of a novel pneumococcal vaccine antigen preferentially expressed during meningitis in mice', *J Clin Invest*, 122: 2208-20.
- Maleki, F., A. Khosravi, A. Nasser, H. Taghinejad, and M. Azizian. 2016. 'Bacterial Heat Shock Protein Activity', *J Clin Diagn Res*, 10: BE01-3.
- Man, W. H., W. A. de Steenhuijsen Piters, and D. Bogaert. 2017. 'The microbiota of the respiratory tract: gatekeeper to respiratory health', *Nat Rev Microbiol*, 15: 259-70.

- Manco, S., F. Hernon, H. Yesilkaya, J. C. Paton, P. W. Andrew, and A. Kadioglu. 2006. 'Pneumococcal neuraminidases A and B both have essential roles during infection of the respiratory tract and sepsis', *Infect Immun*, 74: 4014-20.
- Mandin, P., and J. Johansson. 2020. 'Feeling the heat at the millennium: Thermosensors playing with fire', *Mol Microbiol*, 113: 588-92.
- Mann, B., T. van Opijnen, J. Wang, C. Obert, Y. D. Wang, R. Carter, D. J. McGoldrick, G. Ridout, A. Camilli, E. I. Tuomanen, and J. W. Rosch. 2012. 'Control of virulence by small RNAs in *Streptococcus pneumoniae*', *PLoS Pathog*, 8: e1002788.
- Mardis, E. R. 2017. 'DNA sequencing technologies: 2006-2016', *Nat Protoc*, 12: 213-18.
- Marks, L. R., R. M. Reddinger, and A. P. Hakansson. 2012. 'High levels of genetic recombination during nasopharyngeal carriage and biofilm formation in *Streptococcus pneumoniae*', *mBio*, 3.
- Marra, A., S. Lawson, J. S. Asundi, D. Brigham, and A. E. Hromockyj. 2002. 'In vivo characterization of the *psa* genes from *Streptococcus pneumoniae* in multiple models of infection', *Microbiology (Reading)*, 148: 1483-91.
- Martin, B., C. Granadel, N. Campo, V. Henard, M. Prudhomme, and J. P. Claverys. 2010. 'Expression and maintenance of ComD-ComE, the two-component signal-transduction system that controls competence of *Streptococcus pneumoniae*', *Mol Microbiol*, 75: 1513-28.
- Mascher, T., D. Zahner, M. Merai, N. Balmelle, A. B. de Saizieu, and R. Hakenbeck. 2003. 'The *Streptococcus pneumoniae* *cia* regulon: *CiaR* target sites and transcription profile analysis', *J Bacteriol*, 185: 60-70.
- Masomian, M., Z. Ahmad, L. T. Gew, and C. L. Poh. 2020. 'Development of Next Generation *Streptococcus pneumoniae* Vaccines Conferring Broad Protection', *Vaccines (Basel)*, 8.

- McCool, T. L., T. R. Cate, G. Moy, and J. N. Weiser. 2002. 'The immune response to pneumococcal proteins during experimental human carriage', *J Exp Med*, 195: 359-65.
- McDevitt, C. A., A. D. Ogunniyi, E. Valkov, M. C. Lawrence, B. Kobe, A. G. McEwan, and J. C. Paton. 2011. 'A molecular mechanism for bacterial susceptibility to zinc', *PLoS Pathog*, 7: e1002357.
- Messmer, T. O., J. S. Sampson, A. Stinson, B. Wong, G. M. Carlone, and R. R. Facklam. 2004. 'Comparison of four polymerase chain reaction assays for specificity in the identification of *Streptococcus pneumoniae*', *Diagn Microbiol Infect Dis*, 49: 249-54.
- Miao, X., J. He, L. Zhang, X. Zhao, R. Ge, Q. Y. He, and X. Sun. 2018. 'A Novel Iron Transporter SPD\_1590 in *Streptococcus pneumoniae* Contributing to Bacterial Virulence Properties', *Front Microbiol*, 9: 1624.
- Miller, M. B., and Y. W. Tang. 2009. 'Basic concepts of microarrays and potential applications in clinical microbiology', *Clin Microbiol Rev*, 22: 611-33.
- Minhas, V., R. Aprianto, L. J. McAllister, H. Wang, S. C. David, K. T. McLean, I. Comerford, S. R. McColl, J. C. Paton, J. W. Veening, and C. Trappetti. 2020. 'In vivo dual RNA-seq reveals that neutrophil recruitment underlies differential tissue tropism of *Streptococcus pneumoniae*', *Commun Biol*, 3: 293.
- Mitsi, E., B. Carniel, J. Reine, J. Rylance, S. Zaidi, A. Soares-Schanoski, V. Connor, A. M. Collins, A. Schlitzer, E. Nikolaou, C. Solorzano, S. Pojar, H. Hill, A. D. Hyder-Wright, K. C. Jambo, M. R. Oggioni, M. De Ste Croix, S. B. Gordon, S. P. Jochems, and D. M. Ferreira. 2020. 'Nasal Pneumococcal Density Is Associated with Microaspiration and Heightened Human Alveolar Macrophage Responsiveness to Bacterial Pathogens', *Am J Respir Crit Care Med*, 201: 335-47.
- Moalem, S., E. D. Weinberg, and M. E. Percy. 2004. 'Hemochromatosis and the enigma of misplaced iron: implications for infectious disease and survival', *Biometals*, 17: 135-9.

- Moffitt, K., E. Cheung, T. Yeung, C. Stamoulis, and R. Malley. 2021. 'Analysis of Staphylococcus aureus Transcriptome in Pediatric Soft Tissue Abscesses and Comparison to Murine Infections', *Infect Immun*, 89.
- Molania, R., J. A. Gagnon-Bartsch, A. Dobrovic, and T. P. Speed. 2019. 'A new normalization for Nanostring nCounter gene expression data', *Nucleic Acids Res*, 47: 6073-83.
- Molzen, T. E., P. Burghout, H. J. Bootsma, C. T. Brandt, C. E. van der Gaast-de Jongh, M. J. Eleveld, M. M. Verbeek, N. Frimodt-Moller, C. Ostergaard, and P. W. Hermans. 2011. 'Genome-wide identification of Streptococcus pneumoniae genes essential for bacterial replication during experimental meningitis', *Infect Immun*, 79: 288-97.
- Morimura, A., S. Hamaguchi, Y. Akeda, and K. Tomono. 2021. 'Mechanisms Underlying Pneumococcal Transmission and Factors Influencing Host-Pneumococcus Interaction: A Review', *Front Cell Infect Microbiol*, 11: 639450.
- Morita, M. T., Y. Tanaka, T. S. Kodama, Y. Kyogoku, H. Yanagi, and T. Yura. 1999. 'Translational induction of heat shock transcription factor sigma32: evidence for a built-in RNA thermosensor', *Genes Dev*, 13: 655-65.
- Mraheil, M. A., H. A. Toque, L. La Pietra, J. Hamacher, T. Phanthok, A. Verin, J. Gonzales, Y. Su, D. Fulton, D. C. Eaton, T. Chakraborty, and R. Lucas. 2021. 'Dual Role of Hydrogen Peroxide as an Oxidant in Pneumococcal Pneumonia', *Antioxid Redox Signal*, 34: 962-78.
- Murray, M. J., A. B. Murray, M. B. Murray, and C. J. Murray. 1978. 'The adverse effect of iron repletion on the course of certain infections', *Br Med J*, 2: 1113-5.
- Myoung-Ro, Lee.; Song-Mee, Bae.; Tong-Soo, Kim.; Kwang-Jun, Lee.;. 2006. 'Proteomic Analysis of Protein Expression in Streptococcus pneumoniae in Response to Temperature Shift', *The Journal of Microbiology*, 44: 375-82.

Nanduri, B., P. Shah, M. Ramkumar, E. B. Allen, E. Swiatlo, S. C. Burgess, and M. L. Lawrence. 2008. 'Quantitative analysis of *Streptococcus pneumoniae* TIGR4 response to in vitro iron restriction by 2-D LC ESI MS/MS', *Proteomics*, 8: 2104-14.

Naskulwar, K., and L. Pena-Castillo. 2022. 'sRNARFTarget: a fast machine-learning-based approach for transcriptome-wide sRNA target prediction', *RNA Biol*, 19: 44-54.

Neeleman, C., C. H. Klaassen, D. M. Klomberg, H. A. de Valk, and J. W. Mouton. 2004. 'Pneumolysin is a key factor in misidentification of macrolide-resistant *Streptococcus pneumoniae* and is a putative virulence factor of *S. mitis* and other streptococci', *J Clin Microbiol*, 42: 4355-7.

Nicol, J. W., G. A. Helt, S. G. Blanchard, Jr., A. Raja, and A. E. Loraine. 2009. 'The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets', *Bioinformatics*, 25: 2730-1.

Nicolas, P., U. Mader, E. Dervyn, T. Rochat, A. Leduc, N. Pigeonneau, E. Bidnenko, E. Marchadier, M. Hoebeke, S. Aymerich, D. Becher, P. Bisicchia, E. Botella, O. Delumeau, G. Doherty, E. L. Denham, M. J. Fogg, V. Fromion, A. Goelzer, A. Hansen, E. Hartig, C. R. Harwood, G. Homuth, H. Jarmer, M. Jules, E. Klipp, L. Le Chat, F. Lecoite, P. Lewis, W. Liebermeister, A. March, R. A. Mars, P. Nannapaneni, D. Noone, S. Pohl, B. Rinn, F. Rugheimer, P. K. Sappa, F. Samson, M. Schaffer, B. Schwikowski, L. Steil, J. Stulke, T. Wiegert, K. M. Devine, A. J. Wilkinson, J. M. van Dijl, M. Hecker, U. Volker, P. Bessieres, and P. Noirot. 2012. 'Condition-dependent transcriptome reveals high-level regulatory architecture in *Bacillus subtilis*', *Science*, 335: 1103-6.

Niu, S. Y., B. Liu, Q. Ma, and W. C. Chou. 2019. 'rSeqTU-A Machine-Learning Based R Package for Prediction of Bacterial Transcription Units', *Front Genet*, 10: 374.

Odutola, A., M. O. C. Ota, M. Antonio, E. O. Ogundare, Y. Saidu, E. Foster-Nyarko, P. K. Owiafe, F. Ceesay, A. Worwui, O. T. Idoko, O. Owolabi, A. Bojang, S. Jarju, I. Drammeh, B. Kampmann, B. M. Greenwood, M. Alderson, M. Traskine, N. Devos, S. Schoonbroodt, K. Swinnen, V. Verlant, K. Dobbelaere, and D. Borys. 2017. 'Efficacy of a novel, protein-based



pneumococcal vaccine against nasopharyngeal carriage of *Streptococcus pneumoniae* in infants: A phase 2, randomized, controlled, observer-blind study', *Vaccine*, 35: 2531-42.

Ogunniyi, A. D., K. S. LeMessurier, R. M. Graham, J. M. Watt, D. E. Briles, U. H. Stroehrer, and J. C. Paton. 2007. 'Contributions of pneumolysin, pneumococcal surface protein A (PspA), and PspC to pathogenicity of *Streptococcus pneumoniae* D39 in a mouse model', *Infect Immun*, 75: 1843-51.

Ong, C. L., A. J. Potter, C. Trappetti, M. J. Walker, M. P. Jennings, J. C. Paton, and A. G. McEwan. 2013. 'Interplay between manganese and iron in pneumococcal pathogenesis: role of the orphan response regulator RitR', *Infect Immun*, 81: 421-9.

Ono, S., M. D. Goldberg, T. Olsson, D. Esposito, J. C. Hinton, and J. E. Ladbury. 2005. 'H-NS is a part of a thermally controlled mechanism for bacterial gene regulation', *Biochem J*, 391: 203-13.

Page, A. J., C. A. Cummins, M. Hunt, V. K. Wong, S. Reuter, M. T. Holden, M. Fookes, D. Falush, J. A. Keane, and J. Parkhill. 2015. 'Roary: rapid large-scale prokaryote pan genome analysis', *Bioinformatics*, 31: 3691-3.

Paixao, L., J. Caldas, T. G. Kloosterman, O. P. Kuipers, S. Vinga, and A. R. Neves. 2015. 'Transcriptional and metabolic effects of glucose on *Streptococcus pneumoniae* sugar metabolism', *Front Microbiol*, 6: 1041.

Paixao, L., J. Oliveira, A. Verissimo, S. Vinga, E. C. Lourenco, M. R. Ventura, M. Kjos, J. W. Veening, V. E. Fernandes, P. W. Andrew, H. Yesilkaya, and A. R. Neves. 2015. 'Host glycan sugar-specific pathways in *Streptococcus pneumoniae*: galactose as a key sugar in colonisation and infection [corrected]', *PLoS One*, 10: e0121042.

Pandya, U., C. A. Allen, D. A. Watson, and D. W. Niesel. 2005. 'Global profiling of *Streptococcus pneumoniae* gene expression at different growth temperatures', *Gene*, 360: 45-54.

- Parent du Chatelet, I., Y. Traore, B. D. Gessner, A. Antignac, B. Naccro, B. M. Njanpop-Lafourcade, M. S. Ouedraogo, S. R. Tiendrebeogo, E. Varon, and M. K. Taha. 2005. 'Bacterial meningitis in Burkina Faso: surveillance using field-based polymerase chain reaction testing', *Clin Infect Dis*, 40: 17-25.
- Parker, D., G. Soong, P. Planet, J. Brower, A. J. Ratner, and A. Prince. 2009. 'The NanA neuraminidase of *Streptococcus pneumoniae* is involved in biofilm formation', *Infect Immun*, 77: 3722-30.
- Parsons, J. B., M. W. Frank, C. Subramanian, P. Saenkham, and C. O. Rock. 2011. 'Metabolic basis for the differential susceptibility of Gram-positive pathogens to fatty acid synthesis inhibitors', *Proc Natl Acad Sci U S A*, 108: 15378-83.
- Parui, A. L., V. Mishra, S. Dutta, P. Bhaumik, and K. Bose. 2022. 'Inter-subunit crosstalk via PDZ synergistically governs allosteric activation of proapoptotic HtrA2', *Structure*.
- Pasteur, Louis. 1881. 'Note sur la maladie nouvelle provoquee par la salive d'un enfant mort de la rage', *Bull acad med (Paris)*, 10: 94-103.
- Paterson, G. K., C. E. Blue, and T. J. Mitchell. 2006. 'Role of two-component systems in the virulence of *Streptococcus pneumoniae*', *J Med Microbiol*, 55: 355-63.
- Perez-Sepulveda, B. M., and J. C. D. Hinton. 2018. 'Functional Transcriptomics for Bacterial Gene Detectives', *Microbiol Spectr*, 6.
- Pericone, C. D., S. Park, J. A. Imlay, and J. N. Weiser. 2003. 'Factors contributing to hydrogen peroxide resistance in *Streptococcus pneumoniae* include pyruvate oxidase (SpxB) and avoidance of the toxic effects of the fenton reaction', *J Bacteriol*, 185: 6815-25.
- Perkel, J. M. 2020. 'The software that powers scientific illustration', *Nature*, 582: 137-38.

- Pietrangelo, A. 2004. 'Hereditary hemochromatosis--a new look at an old disease', *N Engl J Med*, 350: 2383-97.
- Pisu, D., L. Huang, J. K. Grenier, and D. G. Russell. 2020. 'Dual RNA-Seq of Mtb-Infected Macrophages In Vivo Reveals Ontologically Distinct Host-Pathogen Interactions', *Cell Rep*, 30: 335-50 e4.
- Plumptre, C. D., C. E. Hughes, R. M. Harvey, B. A. Eijkelkamp, C. A. McDevitt, and J. C. Paton. 2014. 'Overlapping functionality of the Pht proteins in zinc homeostasis of *Streptococcus pneumoniae*', *Infect Immun*, 82: 4315-24.
- Poudel, S., H. Tsunemoto, Y. Seif, A. V. Sastry, R. Szubin, S. Xu, H. Machado, C. A. Olson, A. Anand, J. Pogliano, V. Nizet, and B. O. Palsson. 2020. 'Revealing 29 sets of independently modulated genes in *Staphylococcus aureus*, their regulators, and role in key physiological response', *Proc Natl Acad Sci U S A*.
- Poulsen, L. D., and J. Vinther. 2018. 'RNA-Seq for Bacterial Gene Expression', *Curr Protoc Nucleic Acid Chem*, 73: e55.
- Pozzi, G., L. Masala, F. Iannelli, R. Manganeli, L. S. Havarstein, L. Piccoli, D. Simon, and D. A. Morrison. 1996. 'Competence for genetic transformation in encapsulated strains of *Streptococcus pneumoniae*: two allelic variants of the peptide pheromone', *J Bacteriol*, 178: 6087-90.
- Pracht, D., C. Elm, J. Gerber, S. Bergmann, M. Rohde, M. Seiler, K. S. Kim, H. F. Jenkinson, R. Nau, and S. Hammerschmidt. 2005. 'PavA of *Streptococcus pneumoniae* modulates adherence, invasion, and meningeal inflammation', *Infect Immun*, 73: 2680-9.
- Prudhomme, M., M. Berge, B. Martin, and P. Polard. 2016. 'Pneumococcal Competence Coordination Relies on a Cell-Contact Sensing Mechanism', *PLoS Genet*, 12: e1006113.
- Rajam, G., J. M. Anderton, G. M. Carlone, J. S. Sampson, and E. W. Ades. 2008. 'Pneumococcal surface adhesin A (PsaA): a review', *Crit Rev Microbiol*, 34: 131-42.

Ram, Y., E. Dellus-Gur, M. Bibi, K. Karkare, U. Obolski, M. W. Feldman, T. F. Cooper, J. Berman, and L. Hadany. 2019. 'Predicting microbial growth in a mixed culture from growth curve data', *Proc Natl Acad Sci U S A*, 116: 14698-707.

Ramos-Sevillano, E., G. Ercoli, P. Felgner, R. Ramiro de Assis, R. Nakajima, D. Goldblatt, R. S. Heyderman, S. B. Gordon, D. M. Ferreira, and J. S. Brown. 2021. 'Preclinical Development of Virulence-attenuated *Streptococcus pneumoniae* Strains Able to Enhance Protective Immunity against Pneumococcal Infection', *Am J Respir Crit Care Med*, 203: 1037-41.

Ramos-Sevillano, Elisa, Giuseppe Ercoli, José Afonso Guerra-Assunção, Philip Felgner, Rafael Ramiro de Assis, Rie Nakajima, David Goldblatt, Kevin Kweku Adjei Tetteh, Robert Simon Heyderman, Stephen Brian Gordon, Daniela Mulari Ferreria, and Jeremy Stuart Brown. 2021. 'Protective Effect of Nasal Colonisation with  $\Delta$ cps/piaA and  $\Delta$ cps/proABC*Streptococcus pneumoniae* Strains against Recolonisation and Invasive Infection', *Vaccines*, 9.

Reyes, L. F., M. I. Restrepo, C. A. Hinojosa, N. J. Soni, A. Anzueto, B. L. Babu, N. Gonzalez-Juarbe, A. H. Rodriguez, A. Jimenez, J. D. Chalmers, S. Aliberti, O. Sibila, V. T. Winter, J. J. Coalson, L. D. Giavedoni, C. S. Dela Cruz, G. W. Waterer, M. Witzernath, N. Suttorp, P. H. Dube, and C. J. Orihuela. 2017. 'Severe Pneumococcal Pneumonia Causes Acute Cardiac Toxicity and Subsequent Cardiac Remodeling', *Am J Respir Crit Care Med*, 196: 609-20.

Richmond, C. S., J. D. Glasner, R. Mau, H. Jin, and F. R. Blattner. 1999. 'Genome-wide expression profiling in *Escherichia coli* K-12', *Nucleic Acids Res*, 27: 3821-35.

Ritchie, N. D., T. J. Mitchell, and T. J. Evans. 2012. 'What is different about serotype 1 pneumococci?', *Future Microbiol*, 7: 33-46.

Roche, A. M., S. J. King, and J. N. Weiser. 2007. 'Live attenuated *Streptococcus pneumoniae* strains induce serotype-independent mucosal and systemic protection in mice', *Infect Immun*, 75: 2469-75.

- Romero, P., E. Garcia, and T. J. Mitchell. 2009. 'Development of a prophage typing system and analysis of prophage carriage in *Streptococcus pneumoniae*', *Appl Environ Microbiol*, 75: 1642-9.
- Rosch, J. W., G. Gao, G. Ridout, Y. D. Wang, and E. I. Tuomanen. 2009. 'Role of the manganese efflux system *mntE* for signalling and pathogenesis in *Streptococcus pneumoniae*', *Mol Microbiol*, 72: 12-25.
- Rosenow, C., P. Ryan, J. N. Weiser, S. Johnson, P. Fontan, A. Ortqvist, and H. R. Masure. 1997. 'Contribution of novel choline-binding proteins to adherence, colonization and immunogenicity of *Streptococcus pneumoniae*', *Mol Microbiol*, 25: 819-29.
- Ruckinger, S., R. von Kries, A. Siedler, and M. van der Linden. 2009. 'Association of serotype of *Streptococcus pneumoniae* with risk of severe and fatal outcome', *Pediatr Infect Dis J*, 28: 118-22.
- Ryan, D., L. Jenniches, S. Reichardt, L. Barquist, and A. J. Westermann. 2020. 'A high-resolution transcriptome map identifies small RNA regulation of metabolism in the gut microbe *Bacteroides thetaiotaomicron*', *Nat Commun*, 11: 3557.
- S., Andrews. 2010. 'FastQC: a quality control tool for high throughput sequence data.'
- Samtani, H., G. Unni, and P. Khurana. 2022. 'Microbial Mechanisms of Heat Sensing', *Indian J Microbiol*, 62: 175-86.
- Sanchez, D., M. Boudes, H. van Tilbeurgh, D. Durand, and S. Quevillon-Cheruel. 2015. 'Modeling the ComD/ComE/comcde interaction network using small angle X-ray scattering', *FEBS J*, 282: 1538-53.
- Sanchez-Rosario, Yamil, and Michael D. L. Johnson. 2021. 'Media Matters, Examining Historical and Modern *Streptococcus pneumoniae* Growth Media and the Experiments They Affect', *Frontiers in Cellular and Infection Microbiology*, 11.

Santos-Zavaleta, A., H. Salgado, S. Gama-Castro, M. Sanchez-Perez, L. Gomez-Romero, D. Ledezma-Tejeda, J. S. Garcia-Sotelo, K. Alquicira-Hernandez, L. J. Muniz-Rascado, P. Pena-Loredo, C. Ishida-Gutierrez, D. A. Velazquez-Ramirez, V. Del Moral-Chavez, C. Bonavides-Martinez, C. F. Mendez-Cruz, J. Galagan, and J. Collado-Vides. 2019. 'RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in *E. coli* K-12', *Nucleic Acids Res*, 47: D212-D20.

Sastry, A. V., N. Dillon, A. Anand, S. Poudel, Y. Hefner, S. Xu, R. Szubin, A. M. Feist, V. Nizet, and B. Palsson. 2021. 'Machine Learning of Bacterial Transcriptomes Reveals Responses Underlying Differential Antibiotic Susceptibility', *mSphere*, 6: e0044321.

Sastry, A. V., Y. Gao, R. Szubin, Y. Hefner, S. Xu, D. Kim, K. S. Choudhary, L. Yang, Z. A. King, and B. O. Palsson. 2019. 'The *Escherichia coli* transcriptome mostly consists of independently regulated modules', *Nat Commun*, 10: 5536.

Satzke, C., E. M. Dunne, B. D. Porter, K. P. Klugman, E. K. Mulholland, and group PneuCarriage project. 2015. 'The PneuCarriage Project: A Multi-Centre Comparative Study to Identify the Best Serotyping Methods for Examining Pneumococcal Carriage in Vaccine Evaluation Studies', *PLoS Med*, 12: e1001903; discussion e03.

Schaible, U. E., and S. H. Kaufmann. 2004. 'Iron and microbial infection', *Nat Rev Microbiol*, 2: 946-53.

Schroeder, A., O. Mueller, S. Stocker, R. Salowsky, M. Leiber, M. Gassmann, S. Lightfoot, W. Menzel, M. Granzow, and T. Ragg. 2006. 'The RIN: an RNA integrity number for assigning integrity values to RNA measurements', *BMC Mol Biol*, 7: 3.

Schurch, N. J., P. Schofield, M. Gierlinski, C. Cole, A. Sherstnev, V. Singh, N. Wrobel, K. Gharbi, G. G. Simpson, T. Owen-Hughes, M. Blaxter, and G. J. Barton. 2016. 'How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use?', *RNA*, 22: 839-51.

- Sebert, M. E., L. M. Palmer, M. Rosenberg, and J. N. Weiser. 2002. 'Microarray-based identification of htrA, a *Streptococcus pneumoniae* gene that is regulated by the CiaRH two-component system and contributes to nasopharyngeal colonization', *Infect Immun*, 70: 4059-67.
- Seemann, T. 2014. 'Prokka: rapid prokaryotic genome annotation', *Bioinformatics*, 30: 2068-9.
- Selva, L., R. Benmessaoud, M. Lanaspa, I. Jroundi, C. Moraleda, S. Acacio, M. Inigo, A. Bastiani, M. Monsonis, R. Pallares, Q. Bassat, and C. Munoz-Almagro. 2013. 'Detection of *Streptococcus pneumoniae* and *Haemophilus influenzae* type B by real-time PCR from dried blood spot samples among children with pneumonia: a useful approach for developing countries', *PLoS One*, 8: e76970.
- Seok, J., H. S. Warren, A. G. Cuenca, M. N. Mindrinos, H. V. Baker, W. Xu, D. R. Richards, G. P. McDonald-Smith, H. Gao, L. Hennessy, C. C. Finnerty, C. M. Lopez, S. Honari, E. E. Moore, J. P. Minei, J. Cuschieri, P. E. Bankey, J. L. Johnson, J. Sperry, A. B. Nathens, T. R. Billiar, M. A. West, M. G. Jeschke, M. B. Klein, R. L. Gamelli, N. S. Gibran, B. H. Brownstein, C. Miller-Graziano, S. E. Calvano, P. H. Mason, J. P. Cobb, L. G. Rahme, S. F. Lowry, R. V. Maier, L. L. Moldawer, D. N. Herndon, R. W. Davis, W. Xiao, R. G. Tompkins, Inflammation, and Large Scale Collaborative Research Program Host Response to Injury. 2013. 'Genomic responses in mouse models poorly mimic human inflammatory diseases', *Proc Natl Acad Sci U S A*, 110: 3507-12.
- Shafeeq, S., H. Yesilkaya, T. G. Kloosterman, G. Narayanan, M. Wandel, P. W. Andrew, O. P. Kuipers, and J. A. Morrissey. 2011. 'The cop operon is required for copper homeostasis and contributes to virulence in *Streptococcus pneumoniae*', *Mol Microbiol*, 81: 1255-70.
- Shah, D. H. 2014. 'RNA sequencing reveals differences between the global transcriptomes of *Salmonella enterica* serovar enteritidis strains with high and low pathogenicities', *Appl Environ Microbiol*, 80: 896-906.

- Shak, J. R., H. P. Ludewick, K. E. Howery, F. Sakai, H. Yi, R. M. Harvey, J. C. Paton, K. P. Klugman, and J. E. Vidal. 2013. 'Novel role for the *Streptococcus pneumoniae* toxin pneumolysin in the assembly of biofilms', *mBio*, 4: e00655-13.
- Sharma, C. M., and J. Vogel. 2014. 'Differential RNA-seq: the approach behind and the biological insight gained', *Curr Opin Microbiol*, 19: 97-105.
- Sheldon, J. R., and D. E. Heinrichs. 2015. 'Recent developments in understanding the iron acquisition strategies of gram positive pathogens', *FEMS Microbiol Rev*, 39: 592-630.
- Sheldon, J. R., H. A. Laakso, and D. E. Heinrichs. 2016. 'Iron Acquisition Strategies of Bacterial Pathogens', *Microbiol Spectr*, 4.
- Shi, L., N. Pigeonneau, M. Ventroux, A. Derouiche, V. Bidnenko, I. Mijakovic, and M. F. Noirot-Gros. 2014. 'Protein-tyrosine phosphorylation interaction network in *Bacillus subtilis* reveals new substrates, kinase activators and kinase cross-talk', *Front Microbiol*, 5: 538.
- Sicard, A. M. . 1964. 'A New Synthetic Medium for *Diplococcus Pneumoniae*, and Its Use for the Study of Reciprocal Transformations at the *amiA* Locus', *Genetics*, 50: 31-44.
- Sinha, D., K. Zimmer, T. A. Cameron, D. B. Rusch, M. E. Winkler, and N. R. De Lay. 2019. 'Redefining the Small Regulatory RNA Transcriptome in *Streptococcus pneumoniae* Serotype 2 Strain D39', *J Bacteriol*, 201.
- Skaar, E. P., M. Humayun, T. Bae, K. L. DeBord, and O. Schneewind. 2004. 'Iron-source preference of *Staphylococcus aureus* infections', *Science*, 305: 1626-8.
- Slager, J., R. Aprianto, and J. W. Veening. 2018. 'Deep genome annotation of the opportunistic human pathogen *Streptococcus pneumoniae* D39', *Nucleic Acids Res*.
- . 2019. 'Refining the Pneumococcal Competence Regulon by RNA Sequencing', *J Bacteriol*, 201.



- Smith, A. P., L. C. Lane, T. van Opijnen, S. Woolard, R. Carter, A. Iverson, C. Burnham, P. Vogel, D. Roeber, G. Hochu, M. D. L. Johnson, J. A. McCullers, J. Rosch, and A. M. Smith. 2021. 'Dynamic Pneumococcal Genetic Adaptations Support Bacterial Growth and Inflammation during Coinfection with Influenza', *Infect Immun*, 89: e0002321.
- Spry, C., K. Kirk, and K. J. Saliba. 2008. 'Coenzyme A biosynthesis: an antimicrobial drug target', *FEMS Microbiol Rev*, 32: 56-106.
- Steinmoen, H., E. Knutsen, and L. S. Havarstein. 2002. 'Induction of natural competence in *Streptococcus pneumoniae* triggers lysis and DNA release from a subfraction of the cell population', *Proc Natl Acad Sci U S A*, 99: 7681-6.
- Sternberg, George Miller. 1881. *A fatal form of septicaemia in the rabbit produced by the subcutaneous injection of human saliva: an experimental research* (John Murphy & Company).
- Subramanian, K., B. Henriques-Normark, and S. Normark. 2019. 'Emerging concepts in the pathogenesis of the *Streptococcus pneumoniae*: From nasopharyngeal colonizer to intracellular pathogen', *Cell Microbiol*, 21: e13077.
- Swarthout, T. D., C. Fronterre, J. Lourenco, U. Obolski, A. Gori, N. Bar-Zeev, D. Everett, A. W. Kamng'ona, T. S. Mwalukomo, A. A. Mataya, C. Mwansambo, M. Banda, S. Gupta, P. Diggle, N. French, and R. S. Heyderman. 2020. 'High residual carriage of vaccine-serotype *Streptococcus pneumoniae* after introduction of pneumococcal conjugate vaccine in Malawi', *Nat Commun*, 11: 2222.
- Swarthout, T. D., A. Gori, N. Bar-Zeev, A. W. Kamng'ona, T. S. Mwalukomo, F. Bonomali, R. Nyirenda, C. Brown, J. Msefula, D. Everett, C. Mwansambo, K. Gould, J. Hinds, R. S. Heyderman, and N. French. 2020. 'Evaluation of Pneumococcal Serotyping of Nasopharyngeal-Carriage Isolates by Latex Agglutination, Whole-Genome Sequencing (PneumoCaT), and DNA Microarray in a High-Pneumococcal-Carriage-Prevalence Population in Malawi', *J Clin Microbiol*, 59.

Tai, S. S., C. J. Lee, and R. E. Winter. 1993. 'Hemin utilization is related to virulence of *Streptococcus pneumoniae*', *Infect Immun*, 61: 5401-5.

Takeuchi, N., M. Ohkusu, H. Hishiki, K. Fujii, M. Hotta, S. Murata, and N. Ishiwada. 2020. 'First report on multidrug-resistant non-encapsulated *Streptococcus pneumoniae* isolated from a patient with pneumonia', *J Infect Chemother*.

Tao, H., C. Bausch, C. Richmond, F. R. Blattner, and T. Conway. 1999. 'Functional genomics: expression analysis of *Escherichia coli* growing on minimal and rich media', *J Bacteriol*, 181: 6425-40.

Tettelin, H., K. E. Nelson, I. T. Paulsen, J. A. Eisen, T. D. Read, S. Peterson, J. Heidelberg, R. T. DeBoy, D. H. Haft, R. J. Dodson, A. S. Durkin, M. Gwinn, J. F. Kolonay, W. C. Nelson, J. D. Peterson, L. A. Umayam, O. White, S. L. Salzberg, M. R. Lewis, D. Radune, E. Holtzapple, H. Khouri, A. M. Wolf, T. R. Utterback, C. L. Hansen, L. A. McDonald, T. V. Feldblyum, S. Angiuoli, T. Dickinson, E. K. Hickey, I. E. Holt, B. J. Loftus, F. Yang, H. O. Smith, J. C. Venter, B. A. Dougherty, D. A. Morrison, S. K. Hollingshead, and C. M. Fraser. 2001. 'Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*', *Science*, 293: 498-506.

Thomason, M. K., T. Bischler, S. K. Eisenbart, K. U. Forstner, A. Zhang, A. Herbig, K. Nieselt, C. M. Sharma, and G. Storz. 2015. 'Global transcriptional start site mapping using differential RNA sequencing reveals novel antisense RNAs in *Escherichia coli*', *J Bacteriol*, 197: 18-28.

Todd, E. W., and L. F. Hewitt. 1932. 'A new culture medium for the production of antigenic streptococcal hæmolysin', *The Journal of Pathology and Bacteriology*, 35: 973-74.

Tothpal, A., K. Desobry, S. S. Joshi, A. L. Wyllie, and D. M. Weinberger. 2019. 'Variation of growth characteristics of pneumococcus with environmental conditions', *BMC Microbiol*, 19: 304.

- Troxler, L. J., J. P. Werren, T. O. Schaffner, N. Mostacci, P. Vermathen, M. Vermathen, D. Wuthrich, C. Simillion, S. D. Brugger, R. Bruggmann, L. J. Hathaway, J. Furrer, and M. Hilty. 2019. 'Carbon source regulates polysaccharide capsule biosynthesis in *Streptococcus pneumoniae*', *J Biol Chem*, 294: 17224-38.
- Tsang, H. F., V. W. Xue, S. P. Koh, Y. M. Chiu, L. P. Ng, and S. C. Wong. 2017. 'NanoString, a novel digital color-coded barcode technology: current and future applications in molecular diagnostics', *Expert Rev Mol Diagn*, 17: 95-103.
- Tseng, H. J., A. G. McEwan, J. C. Paton, and M. P. Jennings. 2002. 'Virulence of *Streptococcus pneumoniae*: PsaA mutants are hypersensitive to oxidative stress', *Infect Immun*, 70: 1635-9.
- Turner, A. G., C. Y. Ong, M. J. Walker, K. Y. Djoko, and A. G. McEwan. 2017. 'Transition Metal Homeostasis in *Streptococcus pyogenes* and *Streptococcus pneumoniae*', *Adv Microb Physiol*, 70: 123-91.
- Ulijasz, A. T., D. R. Andes, J. D. Glasner, and B. Weisblum. 2004. 'Regulation of iron transport in *Streptococcus pneumoniae* by RitR, an orphan response regulator', *J Bacteriol*, 186: 8123-36.
- UniProt, Consortium. 2021. 'UniProt: the universal protein knowledgebase in 2021', *Nucleic Acids Res*, 49: D480-D89.
- Usuf, E., C. Bottomley, R. Gladstone, E. Bojang, K. Jawneh, I. Cox, E. Jallow, A. Bojang, B. Greenwood, R. A. Adegbola, S. D. Bentley, P. C. Hill, and A. Roca. 2021. 'Persistent and Emerging Pneumococcal Carriage Serotypes in a Rural Gambian Community After 10 Years of Pneumococcal Conjugate Vaccine Pressure', *Clin Infect Dis*, 73: e3825-e35.
- van Beek, L. F., K. Surmann, H. B. van den Berg van Saparoea, D. Houben, W. S. P. Jong, C. Hentschker, T. H. A. Ederveen, E. Mitsi, D. M. Ferreira, F. van Opzeeland, C. E. van der Gaast-de Jongh, I. Joosten, U. Volker, F. Schmidt, J. Luirink, D. A. Diavatopoulos, and M. I.

- de Jonge. 2020. 'Exploring metal availability in the natural niche of *Streptococcus pneumoniae* to discover potential vaccine antigens', *Virulence*, 11: 1310-28.
- van de Rijn, I., and R. E. Kessler. 1980. 'Growth characteristics of group A streptococci in a new chemically defined medium', *Infect Immun*, 27: 444-8.
- van der Ploeg, J. R. 2010. 'Genome sequence of the temperate bacteriophage PH10 from *Streptococcus oralis*', *Virus Genes*, 41: 450-8.
- van Opijnen, T., and A. Camilli. 2012. 'A fine scale phenotype-genotype virulence map of a bacterial pathogen', *Genome Res*, 22: 2541-51.
- van Vliet, A. H. 2010. 'Next generation sequencing of microbial transcriptomes: challenges and opportunities', *FEMS Microbiol Lett*, 302: 1-7.
- Verhoeven, D., Q. Xu, and M. E. Pichichero. 2014. 'Vaccination with a *Streptococcus pneumoniae* trivalent recombinant PcpA, PhtD and PlyD1 protein vaccine candidate protects against lethal pneumonia in an infant murine model', *Vaccine*, 32: 3205-10.
- Wade, K. R., E. M. Hotze, D. E. Briles, and R. K. Tweten. 2014. 'Mouse, but not human, ApoB-100 lipoprotein cholesterol is a potent innate inhibitor of *Streptococcus pneumoniae* pneumolysin', *PLoS Pathog*, 10: e1004353.
- Wahl, Brian, Katherine L. O'Brien, Adena Greenbaum, Anwesha Majumder, Li Liu, Yue Chu, Ivana Lukšić, Harish Nair, David A. McAllister, Harry Campbell, Igor Rudan, Robert Black, and Maria Deloria Knoll. 2018. 'Burden of *Streptococcus pneumoniae* and *Haemophilus influenzae* type b disease in children in the era of conjugate vaccines: global, regional, and national estimates for 2000–15', *The Lancet Global Health*, 6: e744-e57.
- Wang, H., C. Horbinski, H. Wu, Y. Liu, S. Sheng, J. Liu, H. Weiss, A. J. Stromberg, and C. Wang. 2016. 'NanoStringDiff: a novel statistical method for differential expression analysis based on NanoString nCounter data', *Nucleic Acids Res*, 44: e151.

Wang, P., X. Meng, J. Li, Y. Chen, D. Zhang, H. Zhong, P. Xia, L. Cui, G. Zhu, and H. Wang. 2020. 'Transcriptome profiling of avian pathogenic *Escherichia coli* and the mouse microvascular endothelial cell line bEnd.3 during interaction', *PeerJ*, 8: e9172.

Warrier, I., N. Ram-Mohan, Z. Zhu, A. Hazery, H. Echlin, J. Rosch, M. M. Meyer, and T. van Opijnen. 2018. 'The Transcriptional landscape of *Streptococcus pneumoniae* TIGR4 reveals a complex operon architecture and abundant riboregulation critical for growth and virulence', *PLoS Pathog*, 14: e1007461.

Waters, L. S., and G. Storz. 2009. 'Regulatory RNAs in bacteria', *Cell*, 136: 615-28.

Watkins, E. R., A. Kalizang'Oma, A. Gori, S. Gupta, and R. S. Heyderman. 2022. 'Factors affecting antimicrobial resistance in *Streptococcus pneumoniae* following vaccination introduction', *Trends Microbiol.*

Watson, D. A., D. M. Musher, and J. Verhoef. 1995. 'Pneumococcal virulence factors and host immune responses to them', *Eur J Clin Microbiol Infect Dis*, 14: 479-90.

Wei, S., X. Li, Z. Lu, H. Zhang, X. Ye, Y. Zhou, J. Li, Y. Yan, H. Pei, F. Duan, D. Wang, S. Chen, P. Wang, C. Zhang, L. Shang, Y. Zhou, P. Yan, M. Zhao, J. Huang, R. Bock, Q. Qian, and W. Zhou. 2022. 'A transcriptional regulator that boosts grain yields and shortens the growth duration of rice', *Science*, 377: eabi8455.

Weight, Caroline M., Cristina Venturini, Sherin Pojar, Simon Jochems, Jesús Reiné, Elissavet Nikolaou, Carla Solórzano, Mahdad Noursadeghi, Jeremy S. Brown, Daniela M. Ferreira, and Robert S. Heyderman. 2018. 'Epithelial control of colonisation by *Streptococcus pneumoniae* at the human mucosal surface

Weinberg, E. D. 1999. 'Iron loading and disease surveillance', *Emerg Infect Dis*, 5: 346-52.

Weinberger, D. M., Z. B. Harboe, E. A. Sanders, M. Ndiritu, K. P. Klugman, S. Ruckinger, R. Dagan, R. Adegbola, F. Cutts, H. L. Johnson, K. L. O'Brien, J. A. Scott, and M. Lipsitch. 2010.

'Association of serotype with risk of death due to pneumococcal pneumonia: a meta-analysis', *Clin Infect Dis*, 51: 692-9.

Weinberger, D. M., K. Trzcinski, Y. J. Lu, D. Bogaert, A. Brandes, J. Galagan, P. W. Anderson, R. Malley, and M. Lipsitch. 2009. 'Pneumococcal capsular polysaccharide structure predicts serotype prevalence', *PLoS Pathog*, 5: e1000476.

Weiser, J. N., D. M. Ferreira, and J. C. Paton. 2018. 'Streptococcus pneumoniae: transmission, colonization and invasion', *Nat Rev Microbiol*, 16: 355-67.

Werno, A. M., and D. R. Murdoch. 2008. 'Medical microbiology: laboratory diagnosis of invasive pneumococcal disease', *Clin Infect Dis*, 46: 926-32.

Wessler, S., G. Schneider, and S. Backert. 2017. 'Bacterial serine protease HtrA as a promising new target for antimicrobial therapy?', *Cell Commun Signal*, 15: 4.

Westermann, A. J., K. U. Forstner, F. Amman, L. Barquist, Y. Chao, L. N. Schulte, L. Muller, R. Reinhardt, P. F. Stadler, and J. Vogel. 2016. 'Dual RNA-seq unveils noncoding RNA functions in host-pathogen interactions', *Nature*, 529: 496-501.

Westermann, A. J., S. A. Gorski, and J. Vogel. 2012. 'Dual RNA-seq of pathogen and host', *Nat Rev Microbiol*, 10: 618-30.

Whalan, R. H., S. G. Funnell, L. D. Bowler, M. J. Hudson, A. Robinson, and C. G. Dowson. 2005. 'PiuA and PiaA, iron uptake lipoproteins of Streptococcus pneumoniae, elicit serotype independent antibody responses following human pneumococcal septicaemia', *FEMS Immunol Med Microbiol*, 43: 73-80.

———. 2006. 'Distribution and genetic diversity of the ABC transporter lipoproteins PiuA and PiaA within Streptococcus pneumoniae and related streptococci', *J Bacteriol*, 188: 1031-8.

- WHO. 2017. 'Global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics': 1-7.
- Wick, R. R., L. M. Judd, C. L. Gorrie, and K. E. Holt. 2017. 'Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads', *PLoS Comput Biol*, 13: e1005595.
- Williams, T. M., N. J. Loman, C. Ebruke, D. M. Musher, R. A. Adegbola, M. J. Pallen, G. M. Weinstock, and M. Antonio. 2012. 'Genome analysis of a highly virulent serotype 1 strain of *Streptococcus pneumoniae* from West Africa', *PLoS One*, 7: e26742.
- Wilson, B. R., A. R. Bogdan, M. Miyazawa, K. Hashimoto, and Y. Tsuji. 2016. 'Siderophores in Iron Metabolism: From Mechanism to Therapy Potential', *Trends Mol Med*, 22: 1077-90.
- Wilton, J., P. Acebo, C. Herranz, A. Gomez, and M. Amblar. 2015. 'Small regulatory RNAs in *Streptococcus pneumoniae*: discovery and biological functions', *Front Genet*, 6: 126.
- Wright, A. K., D. M. Ferreira, J. F. Gritzfeld, A. D. Wright, K. Armitage, K. C. Jambo, E. Bate, S. El Batrawy, A. Collins, and S. B. Gordon. 2012. 'Human nasal challenge with *Streptococcus pneumoniae* is immunising in the absence of carriage', *PLoS Pathog*, 8: e1002622.
- Xia, B., J. A. Royall, G. Damera, G. P. Sachdev, and R. D. Cummings. 2005. 'Altered O-glycosylation and sulfation of airway mucins associated with cystic fibrosis', *Glycobiology*, 15: 747-75.
- Xue, R. Y., C. Liu, Q. T. Xiao, S. Sun, Q. M. Zou, and H. B. Li. 2021. 'HtrA family proteases of bacterial pathogens: pros and cons for their therapeutic use', *Clin Microbiol Infect*, 27: 559-64.

- Yang, X. Y., K. He, G. Du, X. Wu, G. Yu, Y. Pan, G. Zhang, X. Sun, and Q. Y. He. 2016. 'Integrated Translatomics with Proteomics to Identify Novel Iron-Transporting Proteins in *Streptococcus pneumoniae*', *Front Microbiol*, 7: 78.
- Yao, J., and C. O. Rock. 2017. 'Exogenous fatty acid metabolism in bacteria', *Biochimie*, 141: 30-39.
- Yesilkaya, H., A. Kadioglu, N. Gingles, J. E. Alexander, T. J. Mitchell, and P. W. Andrew. 2000. 'Role of manganese-containing superoxide dismutase in oxidative stress and virulence of *Streptococcus pneumoniae*', *Infect Immun*, 68: 2819-26.
- Yoo, R., K. Rychel, S. Poudel, T. Al-Bulushi, Y. Yuan, S. Chauhan, C. Lamoureux, B. O. Palsson, and A. Sastry. 2022. 'Machine Learning of All *Mycobacterium tuberculosis* H37Rv RNA-seq Data Reveals a Structured Interplay between Metabolism, Stress Response, and Infection', *mSphere*, 7: e0003322.
- You, Y., C. Fu, X. Zeng, D. Fang, X. Yan, B. Sun, D. Xiao, and J. Zhang. 2008. 'A novel DNA microarray for rapid diagnosis of enteropathogenic bacteria in stool specimens of patients with diarrhea', *J Microbiol Methods*, 75: 566-71.
- Yu, G., L. G. Wang, Y. Han, and Q. Y. He. 2012. 'clusterProfiler: an R package for comparing biological themes among gene clusters', *OMICS*, 16: 284-7.
- Yue, F., Y. Cheng, A. Breschi, J. Vierstra, W. Wu, T. Ryba, R. Sandstrom, Z. Ma, C. Davis, B. D. Pope, Y. Shen, D. D. Pervouchine, S. Djebali, R. E. Thurman, R. Kaul, E. Rynes, A. Kirilusha, G. K. Marinov, B. A. Williams, D. Trout, H. Amrhein, K. Fisher-Aylor, I. Antoshechkin, G. DeSalvo, L. H. See, M. Fastuca, J. Drenkow, C. Zaleski, A. Dobin, P. Prieto, J. Lagarde, G. Bussotti, A. Tanzer, O. Denas, K. Li, M. A. Bender, M. Zhang, R. Byron, M. T. Groudine, D. McCleary, L. Pham, Z. Ye, S. Kuan, L. Edsall, Y. C. Wu, M. D. Rasmussen, M. S. Bansal, M. Kellis, C. A. Keller, C. S. Morrissey, T. Mishra, D. Jain, N. Dogan, R. S. Harris, P. Cayting, T. Kawli, A. P. Boyle, G. Euskirchen, A. Kundaje, S. Lin, Y. Lin, C. Jansen, V. S. Malladi, M. S. Cline, D. T. Erickson, V. M. Kirkup, K. Learned, C. A. Sloan, K. R. Rosenbloom, B. Lacerda de Sousa, K. Beal, M. Pignatelli, P. Flicek, J. Lian, T. Kahveci, D. Lee, W. J. Kent,



- M. Ramalho Santos, J. Herrero, C. Notredame, A. Johnson, S. Vong, K. Lee, D. Bates, F. Neri, M. Diegel, T. Canfield, P. J. Sabo, M. S. Wilken, T. A. Reh, E. Giste, A. Shafer, T. Kutayavin, E. Haugen, D. Dunn, A. P. Reynolds, S. Neph, R. Humbert, R. S. Hansen, M. De Bruijn, L. Selleri, A. Rudensky, S. Josefowicz, R. Samstein, E. E. Eichler, S. H. Orkin, D. Levasseur, T. Papayannopoulou, K. H. Chang, A. Skoultschi, S. Gosh, C. Disteche, P. Treuting, Y. Wang, M. J. Weiss, G. A. Blobel, X. Cao, S. Zhong, T. Wang, P. J. Good, R. F. Lowdon, L. B. Adams, X. Q. Zhou, M. J. Pazin, E. A. Feingold, B. Wold, J. Taylor, A. Mortazavi, S. M. Weissman, J. A. Stamatoyannopoulos, M. P. Snyder, R. Guigo, T. R. Gingeras, D. M. Gilbert, R. C. Hardison, M. A. Beer, B. Ren, and Encode Consortium Mouse. 2014. 'A comparative encyclopedia of DNA elements in the mouse genome', *Nature*, 515: 355-64.
- Zafar, M. A., Y. Wang, S. Hamaguchi, and J. N. Weiser. 2017. 'Host-to-Host Transmission of *Streptococcus pneumoniae* Is Driven by Its Inflammatory Toxin, Pneumolysin', *Cell Host Microbe*, 21: 73-83.
- Zhang, Q., S. X. Xu, H. Wang, W. C. Xu, X. M. Zhang, K. F. Wu, L. Liu, and Y. B. Yin. 2009. 'Contribution of ClpE to virulence of *Streptococcus pneumoniae*', *Can J Microbiol*, 55: 1187-94.
- Zhang, Z., T. B. Clarke, and J. N. Weiser. 2009. 'Cellular effectors mediating Th17-dependent clearance of pneumococcal colonization in mice', *J Clin Invest*, 119: 1899-909.
- Zhou, Z., N. F. Alikhan, K. Mohamed, Y. Fan, Group Agama Study, and M. Achtman. 2020. 'The Enterobase user's guide, with case studies on *Salmonella* transmissions, *Yersinia pestis* phylogeny, and *Escherichia* core genomic diversity', *Genome Res*, 30: 138-52.
- Zhu, A., J. G. Ibrahim, and M. I. Love. 2019. 'Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences', *Bioinformatics*, 35: 2084-92.