Genome-wide analysis of Propionibacterium acnes gene regulation

Yu-fei Lin

Submitted in accordance with the requirement for the degree of PhD

The University of Leeds, Faculty of Biological Sciences

February 2013
Intellectual Property and Publication Statement

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from thesis may be published without proper acknowledgement.

© 2013 The University of Leeds and Yu-fei Lin
Acknowledgements

A thank you to the White Rose Consortium and my initial supervisory team, Dr. Keith Stephenson, Dr. Richard Bojar and Prof. Keith Holland for accepting me up as a PhD student and funded my project.

Thanks to Dr. Jeremy Knapp and Dr. John Heritage for their support during the first year of my PhD, their encouragements have kept me motivated. Thank you to John Wright and Dr. Shuang Guan for training me in the lab, and provided seemingly endless solutions should an experiment goes wrong.

Needless to say much of this work could not have been possible with the guidance and knowledge of my supervisor Dr. Kenny McDowall. The amount of effort he has dedicated to my study is something that I will not forget. The serendipitous working relationship I have developed with him and members of the McDowall lab (David, Justin, Louise and Ayad) have been nothing but joy. There were times where the moods in the lab are shades of grey, however because of them, the hard work became bearable.

Finally, I would like to thank my family for all the support they have provided throughout my entire degree. Your advices and words of wisdom have been invaluable. Also I would also like to thank my girlfriend, Emma Bentham for her understanding and support for the tough time I have put her through during the final stages of my PhD.
Abstract

Sequencing of the genome of *Propionibacterium acnes* produced a catalogue of genes many of which enable this organism to colonise sites in human skin and survive a range of environmental challenges. However as yet, there is little understanding of the relationships and interactions between genes that give rise to an organism, which has major impact on human health and wellbeing as an opportunistic pathogen that causes infections beyond the skin. To provide a platform for better understanding gene regulation in *P. acnes*, this thesis shows using microarrays, reproducible genetic responses to external changes relevant to the skin environment in *P. acnes* can be studied using batch cultures. It then goes on to describe the generation of nucleotide-resolution maps of the primary and secondary transcriptome. The maps were produced by combining differential and global RNA sequencing approaches. Sites of transcriptional initiation, stable RNA processing and mRNA cleavage as well as riboswitches, small non-coding RNAs, vegetative promoters, and previously undetected genes were identified across the genome. In addition, evidence was obtained for the widespread use of leaderless mRNAs, which may be translated by specialised ribosomes. Preliminary evidence for the existence of the latter, in the form of particular ribosomal RNA processing, was obtained. The study also provided statistically robust evidence for pervasive transcription that is associated with both the sense and antisense strands of coding regions. Continuing annotation of the primary and secondary transcriptomes of pathogens will assist comparative and functional genomics approaches and may also aid the modelling of the disease process and therapeutic development.
Table of Contents

Intellectual Property and Publication Statement ................................................................. ii

Acknowledgements ......................................................................................................................... iii

Abstract ........................................................................................................................................... iv

Table of Contents ............................................................................................................................. v

List of Figures .................................................................................................................................. ix

List of Tables ................................................................................................................................... xi

Abbreviations ................................................................................................................................. xii

Chapter 1......................................................................................................................................... 1

1 General introduction ..................................................................................................................... 1

1.1 Socio-economical impact of acne vulgaris ............................................................................. 1

1.2 Pathogenicity of acne vulgaris ............................................................................................... 1

1.2.1 Sebum production and the effect of hormones and stress ................................................. 2

1.2.2 Hyperkeratinisation and comedogenesis ......................................................................... 2

1.2.3 Proliferation of P. acnes - its role in the human skin microbial microflora and in inflammatory response ............................................................... 3

1.2.4 Other contributing factors .............................................................................................. 4

1.3 Treatment for acne .................................................................................................................. 5

1.3.1 Topical agents - benzoyl peroxide, retinoid and isotretinoin ............................................ 5

1.3.2 Antibiotics ......................................................................................................................... 6

1.3.3 Hormonal treatment ......................................................................................................... 7

1.3.4 Alternative treatment ....................................................................................................... 7

1.4 P. acnes associated infection ................................................................................................. 8

1.5 Characterisation of P. acnes growth and the development of a skin-equivalent model 9
1.6 Broad objective and specific aims of thesis ......................................................... 10

Chapter 2 ....................................................................................................................................... 12

2 Establishment of reproducible culture condition and transcriptional response .......... 12

2.1 Introduction .................................................................................................................... 12

2.2 Materials and Methods .................................................................................................. 16

2.2.1 *P. acnes* and its cultivation in synthetic media ....................................................... 16

2.2.2 Generation of *P. acnes* biofilm ................................................................................. 16

2.2.3 Confocal and scanning electron microscopy of biofilm culture ............................. 17

2.2.4 Isolation of bacterial RNA ....................................................................................... 18

2.2.4.1 Kirby Mix .......................................................................................................... 18

2.2.4.2 RiboPure™-Bacteria ......................................................................................... 18

2.2.4.3 Lysozyme and mutanolysin ............................................................................. 18

2.2.5 Reverse-transcription polymerase chain reaction (RT-PCR).................................. 19

2.2.6 Agarose gel electrophoresis .................................................................................... 19

2.3 Results ............................................................................................................................ 20

2.3.1 Reproducibility of batch culture ............................................................................. 20

2.3.2 Effect of potassium down-shift on *P. acnes* growth curve .................................. 21

2.3.3 Continuous culture and biofilm production ........................................................... 24

2.3.4 Quality and yield of RNA isolation ........................................................................ 30

2.3.5 Confirmation of *kdp* operon induction by RT-PCR ................................................ 33

2.4 Discussion ....................................................................................................................... 36

Chapter 3 .................................................................................................................................... 40

3 Analysis of the global transcriptional responses of *P. acnes* to potassium-downshift .... 40

3.1 Introduction .................................................................................................................... 40
3.2 Materials and Methods ........................................................................................................ 42
3.2.1 Gene expression microarray .......................................................................................... 42
3.2.2 Rank Product analysis .................................................................................................. 42
3.2.3 Global RNA sequencing ............................................................................................... 43
3.3 Results ................................................................................................................................ 44
3.3.1 Analysis of the transcriptome for differentially expressed gene using gene expression array and Rank Product algorithm ................................................................................. 44
3.3.2 Analysis of P. acnes transcriptome using global RNA-sequencing ............................... 49
  3.3.2.1 Comparison of gene expression values between global RNA-sequencing and microarray .......................................................................................................................... 50
  3.3.2.2 Analysis of potassium response by global RNA-sequencing .................................. 54
  3.3.2.3 Identification of differentially expressed genes by gRNA-seq ............................... 58
3.4 Discussion ............................................................................................................................. 61

Chapter 4 .................................................................................................................................. 64
4 Primary and secondary transcriptome analysis of P. acnes ................................................... 64
4.1 Introduction ......................................................................................................................... 64
4.2 Materials and Methods ..................................................................................................... 66
  4.2.1 Differential RNA-sequencing ...................................................................................... 66
4.3 Results ................................................................................................................................ 67
  4.3.1 Transcriptional start sites .......................................................................................... 67
  4.3.2 Transcription and maturation of stable RNAs .......................................................... 73
  4.3.3 Vegetative promoters ............................................................................................... 79
  4.3.4 Uncovering multiple layers of regulation .................................................................... 81
  4.3.5 Identification of potential sRNAs ............................................................................... 87
4.3.6 Leaderless mRNAs ................................................................. 89

4.3.7 Re-annotation of protein-coding genes and operon structures ..................... 91

4.4 Discussion .................................................................................. 92

Chapter 5 .......................................................................................... 97

5 Concluding remarks and future work .................................................................. 97

6 Supplementary figures and tables ....................................................................... 102

7 References ......................................................................................... 116
List of Figures

Figure 2.1 Schematic diagram of a basic two-component signal transduction system. .................. 14
Figure 2.2 Growth curves of *P. acnes* static batch culture in different liquid media. ................... 23
Figure 2.3 Nitrocellulose disk model of *P. acnes* biofilm culture. ............................................... 26
Figure 2.4 Confocal laser scanning microscopy of *P. acnes* biofilm. ........................................ 28
Figure 2.5 Scanning electron microscopy of the nitrocellulose substratum and *P. acnes* biofilm. ................................................................. 30
Figure 2.6 Analysis of total nucleic acid isolated using different techniques by gel electrophoresis. ........................................................................................................... 32
Figure 2.7 Analysis of RNA isolated from *P. acnes* grown in liquid with and without potassium downshift and as a biofilm................................................................. 33
Figure 2.8 Analysis of RT-PCR product from RNA isolated with and without potassium downshift. .................................................................................................................... 35
Figure 3.1 M-A scatterplot of the gene expression values from microarray analysis. .................. 46
Figure 3.2. Scatter plot of average expression of protein coding genes from microarray and global RNA-sequencing. ........................................................................................................... 52
Figure 3.3. Global RNA-sequencing data of genes with significant difference in expression levels compared to microarray. ........................................................................................ 54
Figure 3.4. Detection of *P. acnes* gene expression by global RNA-sequencing. ...................... 57
Figure 3.5 M-A scatter plot of global RNA-sequencing reads................................................................................................................................. 59
Figure 3.6. Identification of position showing differential expression by by gRNA-seq in *P. acnes*. ........................................................................................................................................... 60
Figure 4.1. M-A scatterplots of values from differential RNA-seq. ........................................ 68
Figure 4.2. TSSs associated with examples of different classes of RNA. ................................... 72
Figure 4.3. Location of 3' processing sites for tRNA. ............................................................... 74
Figure 4.4. Location of ribosomal RNA processing sites ............................................................ 77
Figure 4.5. Location of RNA processing sites in 5' UTR of *pnp* mRNA. ......................... 78
Figure 4.6. The conserved sequences of promoters associated with the translational machinery. ........................................................................................................................................................................ 80
Figure 4.7. Transcription, promoters and cis-regulatory motifs within the ndr operons. ........ 84
Figure 4.8. RNA-seq analysis of the pqs operon. ................................................................................ 86
Figure 4.9. Examples of P. acnes small RNAs............................................................................. 88
Figure 4.10. Leaderless mRNAs and transcripts of genes requiring reannotation............... 90
Figure S1. Sequence alignment of promoters associated with the translational machinery. ... 114
Figure S2. Degradation of 5'-triphosphorylated RNA using TEX.................................................... 115
List of Tables

Table 3.1 Genes showing significant change in expression with and without potassium downshift. .............................................................................................................................. 47
Table S1. Transcriptional start sites identified for P. acnes..................................................... 102
Table S2. Transcription start sites associated with discrete RNAs that do not cover annotated genes. ........................................................................................................................................ 103
Table S3. P. acnes homologues of genes involved in RNA processing and degradation...... 104
Table S4. List of annotated and novel sRNAs in P. acnes......................................................... 105
Table S5. List of leaderless mRNAs. ..................................................................................... 106
Table S6. List of genes requiring reannotation........................................................................ 109
Table S7. Composition of Holland defined medium............................................................. 110
Abbreviations

Asp – aspartate
ATP – adenosine triphosphate
bp – base pair
CD – cluster of differentiation
cDNA – complementary deoxyribonucleic acid
cm – centimetre
Cy – cyanine
Cys – cysteine
dNA – deoxyribonucleic acid
°C – degree Celsius
dRNA-seq – differential RNA-sequencing
EDTA – ethylenediaminetetraacetic acid
EPS – extracellular signalling peptides
g – standard gravity
Gly – glycine
gRNA-seq – global RNA-sequencing
h – hour
HK – histidine kinase
HSM – Holland Synthetic Medium
IGF – insulin-like growth factor
IL – interleukin
K₂HPO₄ – di-potassium hydrogen phosphate
KEGG – Kyoto Encyclopaedia of Genes and Genomes
KH₂PO₄ – potassium di-hydrogen phosphate
Kₘ – Michaelis constant
Mbp – mega base pair
Met – methionine
min – minute
mL – millilitre
mm – millimetre
mM – millimolar
mRNA – messenger ribonucleic acid
μg – microgram
μL – microlitre
μm – micrometre
μM – micromolar
NaCl – sodium chloride
NAD/NADH – nicotinamide adenine dinucleotide
NCBI – National Centre for Biotechnology Information
ng – nanogram
nm – nanometre
nM – nanomolar
NRPS – non-ribosomal peptide synthetase
nt – nucleotide
OD – optical density
Pa – pascal
PCR – polymerase chain reaction
pfp – probability of false positive
Phe – phenylalanine
RBS – ribosomal binding site
RCA – reinforced clostridial agar
RNA – ribonucleic acid
RR – response regulator
rRNA – ribosomal ribonucleic acid
RT – reverse transcription
s – second
S – Svedberg unit
SDS – sodium dodecyl sulphate
SEM – scanning electron microscopy
sRNA – small ribonucleic acid
TAP – tobacco acid pyrophosphatase
TBE – Tris-borate-EDTA
TEX – Terminator™ S’ phosphate-dependent exonuclease
Thr – threonine
tmRNA – transfer-messenger ribonucleic acid
Tris-Cl – tris(hydroxymethyl)aminomethane chloride
tRNA – transfer ribonucleic acid
TSS – transcriptional start site
TYG – tryptone-yeast extract-glucose
UCSC – University of California, Santa Cruz
UDP – uridine diphosphate
UTR – untranslated region
UV – ultraviolet
v/v – volume/volume
Val – valine
w/v – weight/volume
Chapter 1

1 General introduction

1.1 Socio-economical impact of acne vulgaris

The skin disorder acne vulgaris, commonly known as skin acne, affects all ethnic groups, all ages and both genders. There is a bias towards adolescents, up to 90% of the population from age 15 to 25 will have suffered from acne and more than 30% of sufferers retain scarring from acne lesion, the extent of the scarring depending on the severity of the infection (Zouboulis et al., 2005, Farrar & Ingham, 2004). There is a natural regression of the symptoms from age 27 onwards; however, 5% of sufferers experience chronic acne. Although most dermatological diseases are not life-threatening, the physical symptoms of acne vulgaris can affect greatly the psychologically of sufferers. Depression, anxiety and body-image disorders are significant psychological problems reported by patients. The extreme is suicidal tendencies. Case studies have also linked higher unemployment rate and a lower quality of life to acne vulgaris (Cunliffe, 1986, Law et al., 2010, Simpson, 1993). At the turn of this century, over 5 million visits to general practitioners and over 2 million prescriptions were ascribed per year to the treatment of acne in the US: at an annual cost of $1 billion (Stern, 2000). A more recent review has estimated that the annual cost of treating acne vulgaris in Germany is over €400 million (Radtke et al., 2010). These reports suggest acne vulgaris is more than just a treatable, cosmetically related clinical condition. Further investigation in its management and cost-effective treatment is needed.

1.2 Pathogenicity of acne vulgaris

Despite the report of acne vulgaris dated as far as the 1890s (Mackenzie et al., 1894). The exact aetiology for acne vulgaris is still unclear to date. The current hypothesis for acne vulgaris suggests it is multi-factorial disease that requires the combination of the following factors to
initiate the disease. These factors are increased sebum production, hyperkeratinisation of the follicles and formation of microcomedones, proliferation of P. acnes, and finally the inflammatory response of the host.

1.2.1 Sebum production and the effect of hormones and stress
The majority of acne vulgaris sufferers are the young adolescents, the release of hormone from the pituitary gland and androgen for development during puberty leads to the increase in sebum production (Pochi et al., 1979). The human sebocyte contains receptors for hormones such as testosterone, progesterone and oestrogen, which modulate cell proliferation, sebum production and androgen metabolism (Zouboulis, 2004). Increase in sebum production has also been linked to increase in stress level. Substance P, an eleven amino acid peptide neuropeptide, binds to the neurokinin1 receptor, which has been shown to be involved in regulating emotion, stress and anxiety. It was shown that substance P increases the level of sebum production in human sebocytes by affection hormone production in the hypothalamic-pituitary-adrenal glands (Koo & Smith, 1991, Lee et al., 2008). Other functional receptors found in human sebocyte are for the corticotropin-releasing hormone. The corticotropin-releasing hormone directly induces lipid synthesis in the sebocytes and increases expression of the enzyme Δ5-3β-hydroxysteroid dehydrogenase, which converts dehydroepiandrostosterone to testosterone (Zouboulis, 2009). Another case study has shown a correlation between increased androgens production and males that suffers from chronic acnes (Holland et al., 1998, Marynick et al., 1983). Premenstrual flares of acne in women is further evidence that acne vulgaris has a hormonal basis (Stoll et al., 2001).

1.2.2 Hyperkeratinisation and comedogenesis
The second stage in acne development is the hyperproliferation of keratinocytes in the follicular duct, which leads to desquamation, the shedding of outer layers of skin (Cunliffe et al., 2004). Another consequence of hyperproliferation is obstruction of duct subsequent, which leads to enlargement of the follicle and the formation of microcomedones. Further
accumulation of keratinocytes and sebum causes the destruction of follicle structure and the formation of comedones. The cause of hyperkeratinisation can be of many origins. Inflammatory cytokines have been suggested to be one of the causes (Ingham et al., 1992). Normal human keratinocytes do not produce interleukin-1α, but interleukin-1α is readily isolated from comedones (Walters et al., 1995, Ingham et al., 1998). Another possible cause of hyperkeratinisation is the degradation of sebum by micro-organisms in the follicle lumen to increase the amount of free fatty acid. Free fatty acid has shown to enhance comedone formation in rabbit ear models (Kanaar, 1971, Kligman, 1968). The changes in lipid composition particularly the reduction of linoleic acid, a type of fatty acid, contribute to formation of the comedone (Wertz et al., 1985, Downing et al., 1986). Lower levels of linoleic acid decrease the barrier function of the epidermis and may make the comedone walls more permeable to inflammatory cytokines (Cunliffe et al., 2004). A correlation effect of androgen on linoleic acid concentration has also been determined, as treatment using an anti-androgen cleared the symptoms of acne and returned the linoleic acid level in the skin back to a normal state (Stewart et al., 1986).

1.2.3 Proliferation of P. acnes - its role in the human skin microbial microflora and in inflammatory response

P. acnes is a Gram-positive, microaerophile. It is also non-motile, non-spore forming and pleomorphic in cell morphology, but seen under the microscope as curved rods. Along with other cutaneous propionibacteria, Staphylococcus spp., Corynebacterium spp., and Micrococcus spp., P. acnes comprises the major microflora of human skin, and has a protective role in preventing over proliferation of pathogens (Noble, 1984, Holland et al., 1981). The natural pH of skin ranges from 4.7 to 6.0, the slight acidity inhibits the growth of most pathogen, but does not affect the growth of resident skin flora. The number of P. acnes varies in different regions of the skin, ranging from $10^4$ to $10^6$ per cm$^2$. It is the dominant species found in hair follicles and in other sebum-rich areas of the skin (head, chest and back) (Bojar & Holland, 2004).
P. acnes growth is favoured by the increase in sebum that occurs when follicles become blocked (Toyoda & Morohashi, 2001). Thus, while P. acnes is found on healthy skin naturally, it is consistently found at a higher density in comedones and acne lesions. Whether or not P. acnes is the sole causative agent of acne vulgaris, it appears central to the disease process. P. acnes is thought to have more of a role in the inflammatory response. Neutrophiles have been detected in the ruptured follicles of acne comedone, a discovery that (Kligman, 1974) initiated a different field of investigation in which acne vulgaris is an inflammatory disease rather than a bacterial infection. The degradation of excess sebum causes an increased level of free fatty acid and squalene, a hydrocarbon that is a precursor for steroids, in the follicle duct, which irritates the follicle wall triggering an inflammatory response (Motoyoshi, 1983). P. acnes has been shown to be able to activate different types of innate immune response. Viable P. acnes is able to induce the production of interleukin and tumour necrosis factor-α, the initiation factors for comedogenesis (Graham et al., 2004, Ingham et al., 1992). The destruction of the follicle structure exposes the skin cells to P. acnes, which causes the activation of macrophages through toll-like receptors. This causes the production and release of interleukin-12 and interleukin-8; the former activates natural killer cells while the latter is a chemoattractant for neutrophiles (Kim et al., 2002, Kim, 2005). P. acnes also has a T-cell mitogenic effect and can raise a specific immune response through activation of CD4+ lymphocytes (Jappe, 2003).

1.2.4 Other contributing factors

Researchers have carried out different case studies on other aspects of life that may contribute to the cause of acne vulgaris. Diet has been widely speculated as the second biggest cause of acne with endocrine imbalance as the number one cause. The intake of fatty acid influences the severity of acne symptoms. The intake of omega-6 fatty acid has a pro-inflammatory effect whereas omega-3 fatty acid has an anti-inflammatory effect on the follicles. Omega-3 fatty acid also decreases the level of insulin-like growth factor 1 (IGF-1) (Danby, 2010). IGF-1 stimulates the growth and differentiation of sebocyte; the level of sebum production correlates with IGF-1 level in acne patients (Cappel et al., 2005). Milk and other dairy products contain various animal
hormones including estrogens, progesterone, androgen precursors such as 5α-androstenedione, 5α-pregnanedione, and dihydrotestosterone which have been implicated in comedogenesis. Case studies have shown that there is a positive correlation between the increases in intake in milk for boys during teenage development and the severity of acne through a rise in testosterone level (Adebamowo et al., 2008). Case studies were performed to investigate the association of acne with hereditary factors that might predispose people to the disease. Abnormalities in androgens and lipids in cases of hyperandrogenism have been linked to neonatal acne through the deficiency of 21-hydroxylase, an enzyme involved in the biosynthesis of steroid hormones and stress hormones (Ostlere et al., 1998). A genetic component to acne vulgaris is also suggested by the finding that this disease is inherited by 78% of direct descendant and 75% of the second generation (Wei et al., 2010).

1.3 Treatment for acne

Most treatments used to treat acne are based on drugs that have an anti-comedogenic, anti-inflammatory or anti-microbial activity. Other therapies such as photodynamic therapy and topical agent derived from plant extracts approach the disease in a similar angle, ultimately aimed at reducing the population of microbes or level of inflammation. Equally important is the treatment of psychology factors that accompany acne. The successful treatment of acne takes months and maybe years; therefore, support and consultations are made available for depressed patients. In teenagers, even mild form of acne can cause significant level of distress as a result of comments from peers.

1.3.1 Topical agents – benzoyl peroxide, retinoid and isotretinoin

Benzoyl peroxide is a precursor of free-radical oxygen. Thiol-containing compounds such as cysteine initiate the breakdown of benzoyl peroxide. The free-radical oxygen penetrates the follicles and lowers the microbial count through a bactericidal effect. This broad spectrum agent is applied directly on acne lesion and has shown to be effect against mild acne (Lyons, 1978).
Retinoid is a chemical related to vitamin A, a class of chemicals that influence cellular growth, differentiation and apoptosis. Retinoid has shown to be effective for managing hyperkeratinisation and inflammation of follicles (Rademaker, 2012). Retinoid is a ligand for retinoic-acid and retinoid X receptors. The two receptors normally form a heterodimer that upon binding the ligand enables recruitment of an activator complexed with RNA polymerase to regulate gene expression. Retinoid have been shown to directly or indirectly affect the expression of over 500 genes, the details of which have been documented (Balmer & Blomhoff, 2002, Glass & Rosenfeld, 2000). Isotretinoin, being a derivative form of retinoid (13-cis retinoic acid), significantly reduces the activity and size of sebum glands. Isotretinoin binds to specific retinoic acid response elements in the promoter region of target gene to regulate the transcription of these genes. Isotretinoin reduces inflammation and it has been shown to have an inhibitory effect on the release of lysosomal enzyme from polymorphonuclear leucocytes (Camisa et al., 1982). Lysosomal enzymes contribute to the damage of the follicular wall, which initiates the inflammatory response. Isotretinoin has no direct antimicrobial effect; its ability to control the microbial population likely occurs from the lowering of the follicular sebum concentration through reducing sebocyte activity (Rademaker, 2012). These compounds are normally applied as a topical agent. Oral isotretinoin is only prescribed in severe cases of acne vulgaris as it has been associated with severe depression, photosensitivity, foetal malformation and myalgias have been reported (Jacobs et al., 2001). Topical benzoyl peroxide and isotretinoin are used in conjunction with antibiotics, as they still active against antibiotic-resistant microbes.

1.3.2 Antibiotics
Oral antibiotics are used to treat more severe form of acne, and they work mainly by reducing the microbial count of the follicles, which then indirectly reduces the level of inflammation. Broad spectrum antibiotics against Gram-positive organism are used. These include first-generation tetracyclines (tetracycline and oxytetracycline), second-generation tetracyclines (doxycycline, minocycline and lymecycline) (Simonart et al., 2008), and macrolides (erythromycin,
clindamycin, and azithromycin) (Williams et al., 2012). Both classes of antibiotic act by inhibiting translation, the former (tetracyclines) bind to 30S ribosomal subunit and the latter (macrolides) bind to 50S ribosomal subunit. Antibiotics are not the preferred form of treatment if a topical agent can manage the clinical condition. The broad-spectrum effect of antibiotics disrupts the gut composition of flora and leads to diarrhoea. Pseudomembrane colitis has been reported following the administration of clindamycin (Webster & Graber, 2008). Antibiotic action requires that bacterial cells are actively growing. However, *P. acnes* is a slow-growing organism and the course of treatment can take two months. It has been reported that patients seek relief from symptoms of acne vulgaris (redness or inflammation) in a aesthetic view more than the reduction in the number acne lesions (Jappe, 2003). As the inflammation reduces, the course of antibiotic is often not completed, thereby increasing the likelihood of new antibiotic-resistance *P. acnes* emerging. Cases of tetracycline and macrolide resistance in *P. acnes* have been reported across the globe (Eady et al., 2003, Ross et al., 2003, Eady et al., 2006).

1.3.3 Hormonal treatment

Hormone treatment suppresses or reduces the amount of androgen. This treatment is usually for females that suffer from polycystic ovary syndrome or individuals with hyperandrogenism where they have a raised level of testosterone. Oral contraceptive contains cypoterone acetate and spironolactone, which are progesterone and oestrogen-containing compounds and bind selectively to androgen receptors to reduce their activity. Corticosteroids and inhibitors of 5α-reductase are used to suppress the production of testosterone level. Flutamide is a non-steroid based anti-androgen, however due to its hepatotoxicity its use is limited (Nguyen & Su, 2011).

1.3.4 Alternative treatment

To reduce the effect of hyperkeratinisation, azelic acid is used to peel off the horny layer of the skin. Topical cream containing 20% azelic acid has the equivalent effect of 0.05% tretinoin. This equates to the same level of tretinoin prescribed for managing severe acne management. Thus same result can be achieved without the side effect of the tretinoin steroid. When azelic acid is
used in conjunction with benzoyl peroxide or tetracycline it has shown to be very effect in suppressing the microbial population and keratinisation of follicles. This can however render the skin more sensitive to UV light and experienced burning sensation (Gollnick, 1990). Some patients have also reported reddening of the skin and inflammation (Holland & Bojar, 1993). *P. acnes* naturally produces a pigment, porphyrin, to resist the effect of UV damage to the cell. Photodynamic therapies take advantage of this pigment and utilising laser with blue (415nm) and red (660nm) light, which is absorbed by porphyrin. In combination with benzoyl peroxide, it has shown to effectively reduce *P. acnes* population (Papageorgiou et al., 2000). Side effects of photodynamic therapy are little; patient reported some increase in skin sensitivity and inflammation. Main reason for photodynamic therapy not being routinely used is that it is a relatively expensive treatment. Plant extracts from *Epimedium brevicornum*, *Malus pumila*, *Polygonum cuspidatum*, *Rhodiola crenulata* and *Dolichos lablab* have shown to be effective in eradicating *P. acnes* biofilms. It has also been found that icariin, resveratrol and salidroside are the active compounds in the plant extracts. Further characterisation of these compounds could aid novel drug design (Coenye et al., 2012).

1.4 *P. acnes* associated infection

*P. acnes* has 3 main subtype; type I is associated with acne vulgaris and other dermal associated with the skin. Type II and III are associated with deep tissue infection, nosocomial and post-surgical infection (Brüggemann, 2005). There are increasing numbers of reports stating the recovery of *P. acnes* from surgical sites, often followed a prosthetic joint surgery, neurosurgery, and spinal surgery (Butler-Wu et al., 2011, Dodson et al., 2010, Jakab et al., 1996, McLorinan et al., 2005, Nisbet et al., 2007, Tunney et al., 1999). Prophylactic antibiotic treatment minimises infection by true skin pathogens (*S. aureus*, *Streptococcus pyogenes*) and opportunistic pathogens (*Staphylococcus epidermidis*, *P. aeruginosa*, *P. acnes*). However, as *P. acnes* is slow growing, it can survive a course of antibiotic. As previously mention, reports of antibiotic resistance strain of *P. acnes* have increase in the past 10 years (Coates et al., 2002, Eady et al., 2003). The absence of other microbes, as the course of antibiotic ends, allows *P. acnes* to
successfully establish an infection. *P. acnes* has been recovered from prostheses, often in the form of biofilms (Levy *et al.*, 2008, Ramage *et al.*, 2003), which are known to be recalcitrant to antibiotics (Coenye *et al.*, 2007). The studies of *P. acnes* infection have extended beyond the acne vulgaris. More detail role of *P. acnes* interactions with the host and biofilm growth model in associated with biomaterial have been initiated (Bayston *et al.*, 2007, Tafin *et al.*, 2012).

### 1.5 Characterisation of *P. acnes* growth and the development of a skin-equivalent model

Much work on *P. acnes* physiology has been carried out at the University of Leeds in the laboratory of Prof. Keith Holland. This included the development of a defined synthetic medium, the study of the varying skin-relevant conditions, *e.g.* pH, temperature and oxygen, and the development of a skin-equivalent model.

*P. acnes* was originally cultivated by other labs using complex medium, such as brain heart infusion broth (Freinkel & Shen, 1969). While this allowed biochemical and physiological studies, such as the identification of free fatty produced through extracellular lipase activity as an irritant for skin follicles (Hassing, 1971), the study of cellular factors that induced inflammation was hampered by the complexity of the media. Components of complex media can themselves be antigenic or attract cells of the immune system. This was one of the reasons for developing a synthetic medium for the cultivation of cutaneous propionibacteria. Another reason was that it allowed the identification of factors essential for *P. acnes* growth, *e.g.* biotin, Vitamin B₅ and Vitamin B₆ (Holland *et al.*, 1979). *P. acnes* was found to secrete a range of degradative enzymes, lipases, hyaluronate lyases, proteases and phosphatases, which can degrade the host tissue and are noted as virulence factors (Ingham *et al.*, 1979, Ingham *et al.*, 1980, Ingham *et al.*, 1981, Ingram *et al.*, 1983). As *P. acnes* is normally resident in the skin microflora, Holland et al. aimed to determine the changes in specific skin-relevant variables or the combination of them that leads to increased activities of these degradative enzymes, and so changes *P. acnes* from being a benign part of the flora into a pathogen. The effect of varying pH,
oxygen, nutrient availability, on the growth rate, biomass and the secreted exoenzyme of *P. acnes* were investigated (Cove *et al.*, 1983, Greenman & Holland, 1985, Greenman *et al.*, 1981, Greenman *et al.*, 1983).

One condition that of the follicle that cannot be reproduced using liquid cultures is the high lipid, low water content. This prompted the development of a human skin equivalent model in which a dermal matrix of fibrin containing fibroblasts is seeded with human keratinocyte to generate a stratified epidermis (Holland *et al.*, 2008). This skin model has been shown to support the colonisation of skin flora, *S. epidermidis, P. acnes* and *Malassezia spp.* The same skin equivalent model was used to assess the differences in gene expression of the keratinocytes using two-channel microarray from colonisation of a skin pathogen *S. aureus*. The study showed that there was an upregulation of diverse gene involved in the innate immune response including toll-like receptor 2, β defensin 4, peptidoglycan recognition proteins; proinflammatory cytokines including interleukins IL-1β, IL-1α, IL-17C, IL-20, IL-23A, tumour necrosis factor and lymphotixin β (Holland *et al.*, 2009).

### 1.6 Broad objective and specific aims of thesis

As indicated in the previous sections, much of the published work on *P. acnes* describes clinical infections and the associated immunology, the rise of antibiotic resistance in hospital isolates and the development of new or improved therapies. In addition to contributing to all of these areas, research at the University of Leeds has established defined conditions for the reliable culture of *P. acnes* in the laboratory, including in the description of skin models. However, it was the successful use of microarrays to identify the innate response of keratinocytes to colonisation by *S. aureus* (Holland *et al.*, 2009) in particular that provided the catalyst for the work undertaken during this thesis. It demonstrated to those involved that much could be learned from analysing transcriptome data. The next logical step was to establish a platform for analysing the transcriptome of *P. acnes*, which the long-term view of being able study alongside equivalent data for colonised keratinocyte. The genome sequence of a clinical isolate of *P.*
Acnes had been determined and annotated by others to produce an initial catalogue of genes (Brzuszkiewicz et al., 2011, Horváth et al., 2012, Hunyadkúrti et al., 2011, Vörös et al., 2012).

The broad objective of this thesis was to utilise this expertise in culturing clinical isolates of P. acnes, to establish a platform for functional genomics, which can be defined as the investigation of the function of genes (and their products), revealed by genome sequencing, use high-throughput rather than more traditional 'gene-by-gene' methods. As described in the chapters that follow, a transcriptome platform has been established that can now be extended to investigate P. acnes growth not only in the skin equivalent model, but actual follicles.

The specific aims were to:

1. establish the reproducibility of culturing P. acnes and induce a stress(es) that would produce a genetic response(s) that was at least in part predictable;

2. confirm using microarray technology that the response was reproducible and describe the nature of the overall response; and

3. map the primary and secondary transcriptomes of P. acne using deep RNA sequencing approaches, thereby improving our understanding of gene structure, and the underlying mechanisms that control gene expression.
Chapter 2

2 Establishment of reproducible culture condition and transcriptional response

2.1 Introduction

This chapter describes fundamental steps towards establishing an experimental platform from which to study genome-wide responses of *P. acnes*, at the level of transcription, to challenges that this organism is likely to face within the skin environment. The first step was to obtain reproducible growth profiles. Two cultivation methods are used routinely in bacteriology, batch and continuous culture (Wanner & Egli, 1990, Novick, 1955, Monod, 1949, Harder & Kuenen, 1977). The former is a closed system, in which bacteria are supplied with a fixed amount of nutrient. This produces distinct phases that are temporally separated; lag, exponential growth, stationary and eventually death. During the lag phase, bacteria adapt to their surrounding environment, synthesising the necessary RNA and protein for reproduction, often exhibiting little or no growth. The exponential phase represents the doubling of the bacteria at their maximum rate for a given condition and appears as a straight line with a positive gradient when log values of cell numbers are plotted against time. In stationary phase, the growth of bacteria slows or halts due to a decline in the amount of available nutrients or the production of inhibitory secondary metabolites. The transition from stationary into death phase is often not clear from measurements of optical density as bacteria can persist in a viable but non-culturable state (Oliver, 2005). In contrast, continuous culture is an open system. It consists of a central culturing vessel into which media is pumped at a specific rate from a reservoir. This rate of nutrient feed controls the rate of growth. Culture volume is maintained by letting the excess volume flow out. This movement of liquid removes secondary metabolites or other elements that could affect growth of the culture. Variables such as atmosphere, pH
and temperature can also be monitored and kept constant, which in turn allows the culture to be maintained in a constant state (Brogden & Guthmiller, 2002).

Planktonic culture although well established and commonly used in labs, rarely represents the bacteria in their natural state. It is proposed that some bacteria exist as single- or multi-species microcolonies surrounded by an extracellular matrix. This mode of growth model provides higher tolerance to environmental and chemical stresses (Hall-Stoodley et al., 2004). Increasingly, bacteria are being cultivated to produce biofilms, which are typically characterised by slow growing cells that are firmly attached to each other and a substratum (Costerton et al., 1995, Lindsay & von Holy, 2006). *P. acnes* has shown ability to adhere to different surfaces; published methods of producing *P. acnes* biofilms involve cultivation on prosthetic biomaterials such as silicon, steel and titanium or using flat-bottom 96-well microtitre plates (Ramage et al., 2003, Bayston et al., 2007, Coenye et al., 2007). The development of a simple biofilm model is also described in this chapter. The cell composition of biofilms is expected to be heterogeneous relative to continuous cultures or batch cultures during exponential growth (Hall-Stoodley et al., 2004, Stewart & Franklin, 2008).

The second aim was to ensure that culture conditions could be altered to produce reproducible changes in the physiology of cells as measured at the level of transcription. *P. acnes* contains homologues of the KdpDE two-component signal transduction system, which in other bacteria has been shown to be involved in potassium sensing and osmoprotection (Csonka & Hanson, 1991, Cholo et al., 2009, Ballal & Apte, 2005). Two-component signal transduction system is one of the many mechanisms in which bacteria sense and adapt to the environment by regulation of gene expression (Capra & Laub, 2012). A basic two-component system is composed of a sensor kinase and its cognate response regulator. More complex signal transduction systems are referred as phosphorelay systems (Fabret et al., 1999, Hoch, 2000), which include additional sensor kinases and response regulators to provide more stringent regulation of gene expression. The regulation of sporulation in *Bacillus spp.* is one such example (Molle et al., 2003). Sensor kinase, often a membrane-associated protein, upon receiving a
specific stimulus in the N-terminal signal input domain initiates the autophosphorylation of the conserved histidine residue by the autokinase domain in the C-terminal end. The phosphoryl group is transferred to the conserved aspartate residue on its cognate response regulator. The phosphotransfer induces a conformational change to the response regulator and disrupts the interaction between the receiver domain and DNA-binding domain (Figure 2.1).

Figure 2.1 Schematic diagram of a basic two-component signal transduction system. The signal input domain is mostly, but not exclusively, associated with the cell membrane. The phosphotransfer and ATP-binding sub-domain of autokinase domain catalyse the autophosphorylation of conserved histidine residue. The phosphoryl group is transferred to conserved aspartate residue on the receiver domain of the response regulator, which then activates the DNA-binding domain for binding of target genes.

The DNA-binding domain allows the response regulator to act as a transcription factor. When the response is no longer required, the sensor kinase can also exhibit phosphatase activity to dephosphorylate the response regulator to negate the response (Hsing et al., 1998, Stock et al.,
The gene encoding the sensor kinase and its cognate response regulator are generally adjacent to each other and are transcribed from the same promoter site. The phosphorylated response regulator can often exhibit an autoregulatory characteristic by binding to its promoter site and recruiting RNA polymerase (Hoch et al., 1995, Parkinson, 1993).

The KdpDE two-component system is widely conserved in bacteria (Frymier et al., 1997, Ballal & Apte, 2005, Xue et al., 2011, Cholo et al., 2009). It has been most studied in *E. coli*, but also in *Salmonella typhimurium*, *Mycobacterium tuberculosis*, *Anabaena sp.* and *Staphylococcus aureus*. Maintenance of intracellular potassium concentration is important for regulation of pH and activity of enzymes (Sueltor, 1970). In a hypotonic environment (<2 mM potassium), the sensor kinase KdpD autophosphorylates the conserved histidine residue (Jung et al., 2001), and this subsequently leads to the transfer of phosphate to the aspartate residue on the KdpE response regulator. The activated KdpE dimerises and binds to the promoter region upstream of the *kdpFABC* and upregulates its expression. The *kdpFABC* encodes for the Kdp-ATPase structural proteins (Altendorf et al., 1994). KdpA positioned on the outer surface of membrane has a high affinity ($K_m$ of 2 μM) for potassium ions, KdpB is the P-type ATPase for translocating the cation, KdpC is for assembly of the ATPase and KdpF is needed for protein stability but is not essential (Epstein, 2003, Ballal et al., 2007). KdpDE is found constitutively expressed at low level. Under potassium limiting conditions, the expression of *kdpFABC* can increase in *E. coli* by 1000 fold (Hamann et al., 2008). The reproducibility of the biological replicates generated in this investigation will be determined by using this predictable response as a control.
2.2 Materials and Methods

2.2.1 P. acnes and its cultivation in synthetic media

Propionibacterium acnes strain KPA171202 was obtained from Ulm University, Göttingen, Germany (Bruggemann et al., 2004) and cultivated in an anaerobic workstation (MACS-MG-1000, Don Whitely Scientific) at 34°C under 80% [v/v] N₂, 10% [v/v] CO₂, and 10% [v/v] H₂. All analyses were done using cells cultivated without shaking in 100 mL of modified Holland Synthetic Medium (HSM; Holland et al., 1979) in a 250 mL Erlenmeyer flask. Inocula were prepared in two stages. First a single colony isolated from the surface of reinforced clostridial agar (RCA) (Farrar et al., 2007) was used to inoculate 10 mL of TYG broth (1.0% [w/v] tryptone, 0.5% [w/v] yeast extract, 0.25% [w/v] D-glucose) in 30 mL plastic universals. After growing to stationary phase, an aliquot was used to inoculate 100 mL of TYG broth to an OD₆₀₀ of 0.2. The culture was then incubated to an OD₆₀₀ of 1.0, after which cells were harvested by centrifugation (3,000 x g for 20 min) and washed by resuspending in 10 mL of HSM pre-warmed to 34°C and then harvest was repeated. Finally, the cells were resuspended in 10 mL of pre-warmed HSM and an appropriate aliquot was used to inoculate 100 mL of pre-warmed HSM to an OD₆₀₀ of 0.2. To study the effects of a potassium down-shift, a 100 mL culture of P. acnes was prepared as previously described and grown to an OD₆₀₀ of 1.0, after which the culture was separated into two equal halves and cells were harvested as described above. One half was washed using standard HSM, used to inoculate 100 mL of fresh HSM and reincubated. The other half was processed in the same way, except using HSM without potassium di-hydrogen phosphate and di-potassium hydrogen phosphate. After 1 h of reincubation, 12.5 mL of STOP solution (95% [v/v] ethanol; 5% [v/v] phenol, (Lin-Chao & Cohen, 1991) was added to inhibit cell metabolism, and the cells were harvested by centrifugation. When necessary, cell pellets were stored frozen at -80°C.

2.2.2 Generation of P. acnes biofilm

The technique was adapted from the cellulose disk model described in the thesis of Dr. Victoria Ryder (Ryder, 2010). Inoculum for biofilm was prepared as above. Sterile nitrocellulose filter
disk (25 mm diameter, 0.22 μm pore size, Millipore) was used as a substratum. A filter disc was steeped in a 10 mL liquid culture in TYG broth, prepared as described above. Inoculated discs were placed on RCA and incubated anaerobically at 34°C for 7 days. Non-adherent cells were removed by washing the disc in sterile saline (0.9% [w/v] NaCl) then return to incubation on fresh RCA. The process was repeated for 4 weeks or until sufficient growth was obtained. Upon harvesting the biofilm, non-adherent cells were removed by the washing step, disc was transferred to a Petri-dish containing 10 mL of sterile saline and 1.25 mL STOP solution. Attached cells were removed using a sterile spatula and transferred to a 25 mL Falcon tube. Cells were then harvested by centrifugation and pellets were stored at -80°C.

2.2.3 Confocal and scanning electron microscopy of biofilm culture
Biofilms were visualised using confocal laser scanning microscopy and low-temperature scanning electron microscopy (SEM), both of which were performed by Jackie Hudson, Leeds Dental Institute. The nitrocellulose disk was cut to approximately 1.0 cm² with a scalpel. A sample was stained with LIVE/DEAD® BacLight™ Bacterial Viability kit (Invitrogen) to assess the proportion of viable cells in the biofilm and visualised under a 40 x wet mount with confocal laser scanning microscopy (Leica TCS SP2). The specimen was scanned for emission of the fluorescent dyes from 500 to 700 nm. Images were captured and collated in 10 μm increments measured from the base of specimen until 500 μm was reached.

The biofilm was air-dried in a sterile Petri-dish for 48 h before visualising with low-temperature SEM (Hitachi S-3400N). A filter disc fragment was placed on a small pedestal with an adhesive platform and wetted with sterile water. The sample was placed on the stage and the chamber was set to atmospheric pressure of 50 Pa and at -20°C.
2.2.4 Isolation of bacterial RNA

2.2.4.1 Kirby Mix

Cell pellets of *P. acnes* were resuspended in Kirby mix (Kieser et al., 2000), 1.0 OD_{600} unit of cells per 100 μL of mix, and then transferred to Lysing Matrix B tubes containing fine silica beads (MP Biomedical). Tubes were then placed in high-speed benchtop homogenizer (FastPrep®-24, MP Biomedical; set at 6.5M/s). Cells were lysed by three cycles of homogenizing for 1 min followed by cooling in an ice-water bath for 1 min. Nucleic acid was extracted using an equal volume of acidic phenol: chloroform: isoamyl alcohol (50: 50: 1), and then chloroform: isoamyl alcohol (49: 1). Nucleic acid in the aqueous phase was precipitated by adding NaCl to 150 mM and 2.5 x volumes of 100% [v/v] ethanol, then chilling at -20°C for 1 h, and finally harvesting by centrifugation (13,000 x g) for 30 min at 4°C. Nucleic acid pellets were washed twice with 700 μL of 70% [v/v] ethanol, air dried for 5 min and resuspended in RNase-free water. To remove contaminating DNA, samples were treated with DNase I using conditions described by the vendor (Ambion) and extracted with phenol: chloroform as described above. The concentration and integrity of RNA samples was determined using a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific) and gel electrophoresis (Sambrook & Russell, 2006), respectively.

2.2.4.2 RiboPure™-Bacteria

Protocol for RNA extraction using this commercial kit was carried out with a single modification to manufacturer’s instruction; the cells were homogenised using the FastPrep®-24 benchtop homogenizer instead of the vortex adaptor (Ambion P/N AM10024) suggested by the manufacturer.

2.2.4.3 Lysozyme and mutanolysin

Cell pellet of *P. acnes* was resuspended in 180 μL of enzymatic lysis buffer (20 mM Tris-Cl pH_8; 2 mM EDTA; 1.2% [v/v] Triton X-100) and added to the reaction mix, 20 μL of either lysozyme (200 ng/μL) or mutanolysin (6.25 unit/μL). The suspension was incubated at 37°C for 1 h, with a gentle inversion after a 30 min interval. Then, 20 μL of 10% [w/v] SDS was added and the
suspension was mixed by vortexing. Equal volume of phenol was added to the suspension before being transferred into a Lysing Matrix B tube. The remainder of the extraction protocol are as described in Kirby Mix extraction.

2.2.5 Reverse-transcription polymerase chain reaction (RT-PCR)

Complementary DNA was synthesised using SuperScript® RT III (Invitrogen) with random hexamer (100 nM) and 200 ng of RNA template, the rest of the protocol were carried out as stated by manufacturer with no modification. The cDNA was diluted with RNase-free water to 100 μL. Primer pairs were designed for PPA0010 and PPA0015 using Primer3 (http://frodo.wi.mit.edu/), and ordered from Sigma-Aldrich. The sequences for the primers are as followed: PPA0010F: 5’-CCCGTACTGGTCAGCGTTTA-3’; PPA0010R: 5’-GCCGTCTGCTGTACAGGT-3’; PPA0116F: 5’-CGGCAAGCAACTACTCATCA-3’; PPA0116R: 5’-TAAAGATGATCGCCGAGAGC-3’. The PCR reaction was carried out using GoTaq® DNA polymerase (Promega). The PCR master mix was prepared according to manufacturer’s instruction with a final reaction volume of 25 μL. Master mix was aliquoted to 0.2 mL tubes and the 2 μL of the diluted cDNA template was added last. Tubes were briefly vortexed and the mix was pooled by brief centrifugation to collect the reaction mixture before placing in the thermal cycler (Techne). The program for the thermal cycler was as follows: initial denaturation of 95°C for 5 min and then 30 cycles of 95°C for 30 s, 60°C for 30 s, and 72°C for 30 s, before a final extension at 72°C for 5 min. Finished cycle were held at 4°C.

2.2.6 Agarose gel electrophoresis

The amount of agarose (Melford) required for the desired weight to volume ratio were measured into a 100 mL Erlenmeyer flask and suspended in 30 mL of 1 x TBE (Severn Biotech) and 3 μL of 10,000 x SYBR-safe nucleic acid stain (Invitrogen). Agarose was melted by heating with microwave at highest power in 20 s intervals until fully dissolved. Gel was then casted and placed in an electrophoresis tank (Bio-Rad) topped with 1 x TBE running buffer. RNA samples were mixed with 2 x RNA loading dye (Ambion) before loaded onto the gel.
2.3 Results

2.3.1 Reproducibility of batch culture

Prior to the work described herein, empirical studies by the laboratory of Prof. Keith Holland had established a synthetic medium in which *P. acnes* grows well (Holland et al., 1979). The standard inoculums for such cultures were cells grown in a rich broth containing tryptone-yeast extract-glucose (TYG; (Farrar et al., 2007). Moreover, it was reported that the cells for inocula must be isolated during the exponential-growth phase to produce reproducible growth in Holland synthetic medium (HSM; Table S7; Keith Holland, pers. comm.). Thus, initial experiments focussed on defining growth profiles for *P. acnes* in static batch cultures of TYG and subsequently HSM (Figure 2.2). Strain KPA171202 was used for these studies as at the time it was the only *P. acnes* for which a genome sequence had been determined (Bruggemann et al., 2004), and because it is a clinical isolate of type I *P. acnes*, recovered from an acne lesion and post-surgical infections.

The growth of *P. acnes* in TYG broth was typical of batch culture, a short lag phase was followed by an exponential growth phase and then a stationary phase during which growth slowed and eventually stopped (Figure 2.2, panel A). The doubling-time and specific growth rate during exponential growth in TYG broth were 5.8 h and 0.119 h⁻¹, respectively. Cells were isolated when the culture reached an OD₆₀₀ of ~1.0 to produce inoculums for HSM. At this point a significant proportion of the maximum achievable biomass had been produced and the cells were still growing exponentially.

Cells passaged into HSM appeared to grow exponentially without a discernible lag phase (Figure 2.2, panel B). The doubling-time and specific growth rate during exponential growth in HSM were 6.2 h and 0.111 h⁻¹, respectively. These values are similar to those obtained above for cells grown in TYG broth. There are no published values on the maximum specific growth rate for the *P. acnes* strain in the medium used in this study. Characterisation of *P. acnes* growth physiology carried out by Holland & Greenman was on *P. acnes* strain P37 using continuous
culture (Greenman et al., 1981, Greenman et al., 1983, Cove et al., 1983, Greenman & Holland, 1985). Strain P37 is also a clinical isolate of type-I *P. acnes*. Holland & Greenman cultivated *P. acnes* P37 in semi-synthetic medium; tryptone, supplemented with vitamins and mineral salts. The maximum specific growth rate obtained was 0.21 h⁻¹. The exponential phase in HSM was about 8 h longer than in TYG; the cells reaching a much higher final OD₆₀₀. Cells cultured in HSM had a tendency to clump and settle in the bottom of the culture flask. Therefore, for OD₆₀₀ analysis the culture flask was swirled vigorously and withdrawn aliquots were vortexed thoroughly.

2.3.2 Effect of potassium down-shift on *P. acnes* growth curve

For cultures grown in HSM, the exponential phase exhibited slower growth after an OD₆₀₀ of 2.0. Thus, an OD₆₀₀ of 1.0 was judged to be mid-exponential and used for subsequent molecular analyses. Growth in exponential phase in HSM was monitored and a linear relationship between Log₁₀(OD₆₀₀) and time was observed (data not shown). During the course of these experiments, the growth of *P. acnes* following potassium downshift was also investigated by inoculating cells into HSM to which potassium di-hydrogen phosphate (KH₂PO₄) and di-potassium hydrogen phosphate (K₂HPO₄) had not been added (Figure 2.2, panel B). There was an initial concern that the culture would not grow after the downshift (Holland, pers. comm.). However, this proved not to be the case and the culture grew, albeit more slowly. There was no distinct exponential phase following the potassium downshift. Cells grew without a discernible lag phase, but their growth rate appeared to decrease steadily with time. It was concluded that the decreasing growth rate reflected the utilisation of phosphate reserves that were accumulated during growth on TYG. Inorganic polyphosphate have been found in cell to serve as an energy and phosphate store to resist environmental stress or starvation. From the annotated *P. acnes* genome, a gene cluster PPA0338 to PPA0340 was found, which encodes for the Pst phosphate transport permease and PPA0341, a phosphate-binding protein. Directly upstream of these genes, PPA0343 encodes polyphosphate kinase, which catalyses the conversion of terminal gamma-phosphate of ATP to polyphosphate (Gadd, 1990, Kornberg et
At the midpoint in the culture, the doubling-time and specific growth rate during exponential growth in HSM were 6.2 h and 0.111 h⁻¹, respectively. The growth rate is at least 8-fold slower after the potassium downshift. Moreover, in this medium cell clumping was more apparent. Indeed, cells formed a structure resembling a biofilm, a thin film that stuck to the bottom on the flask. As described above, this mode of growth may be induced to allow *P. acnes* to survive under stressful conditions.
Figure 2.2 Growth curves of *P. acnes* static batch culture in different liquid media. Panel A shows the average OD$_{600}$ of 4 independent growth curves of *P. acnes* in TYG broth. Panel B shows the average OD$_{600}$ reading of 5 independent *P. acnes* in HSM with and without added potassium, which are represented black and red points, respectively. For all panels, the error bars indicate standard deviation of the OD$_{600}$ readings. The estimated time for the entry and exit of exponential phase are marked by the dashed vertical lines.
2.3.3 Continuous culture and biofilm production

Cells harvested from cultures grown in batch using TYG broth, were also used to inoculate continuous culture vessel. Temperature, oxygen, and pH probes for the chemostats were initially calibrated with water. The unit was autoclaved and water was pumped out with pressurised gas and replaced with TYG or HSM. Cells from batch culture was harvested and used as an inoculum. Cells were resuspended in a small volume of the same medium as used in the chemostat and injected into the culture vessel with a needle. While it was relatively straightforward to establish a viable culture and then growth at a specific rate using TYG or HSM, similar attempts to establish a continuous culture using HSM in the absence of potassium phosphate proved unsuccessful (data not shown). This was due not only to the long doubling time, which was estimated to exceed 24 h, but the strong tendency of cells to form large clumps (approximately 2.0 mm) rather than disperse uniformly, even with the operation of the impeller. Therefore, the decision was taken to drop continuous culture in favour of batch.

In response to the tendency of *P. acnes* cells to clump during the initial experiments described above and the increasing amount of evidence that biofilms are important in colonisation and pathogenicity (Jahns et al., 2012), it was decided that biofilms should be included in this study. The generation of *P. acnes* biofilm has been published (Stepanovic et al., 2000, Holmberg et al., 2009, Coenye et al., 2007). The technique produces a submerged biofilm on the wells of a microtitre plate; stationary phase culture of *P. acnes* grown in a rich broth (brain heart infusion medium/reinforced clostridial medium) was used to inoculate wells of a microtitre plate containing the culture medium. After 48 h incubation, culture broth was aspirated and non-adhering cells where removed by washing the culture with sterile saline with agitation. This protocol would not yield sufficient biomass to isolate the amount of RNA needed for transcriptomic studies. Therefore, a mode of culture developed by Dr. Victoria Ryder (Leeds) for *S. aureus* was adopted in which cells are cultivated on nitrocellulose discs (25 mm) as a substratum and placed on solid medium. Discs were inoculated using an aliquot of cells from a liquid culture in TYG broth and then transferred onto reinforced clostridial agar where they were incubated for minimum of 7 days. As shown in Figure 2.3, *P. acnes* adhered to the
nitrocellulose disc after washing with saline. Detachment of the *P. acnes* biofilm by degrading the extracellular matrix to harvest all the adherent cells was attempted (Figure 2.3, panel D). As the composition of the *P. acnes* biofilm extracellular matrix is unknown, a detachment method for *Staphylococcus aureus* and *S. epidermidis* biofilm was tested, using cellulase and sodium metaperiodate, respectively (Ryder, 2010). This technique however proved to be unsuccessful. Biofilm culture was liberated by manually scraping the surface of the disc. The biofilm was cultivated for a further three months, where a visible film like structure formed (Figure 2.3, panel F).
Figure 2.3 Nitrocellulose disc model of *P. acnes* biofilm culture. A nitrocellulose disc was inoculated with *P. acnes* grown in TYG broth as previously described. Panel A shows the biofilm grown on RCA after 1 week. Panel B shows the removal of non-adherent cells by saline washes. Panel C shows the attempt of detachment using cellulase. Panel D shows the detachment of biofilm material by manual scraping. Panel E shows the growth of *P. acnes* biofilm after 4 weeks. Panel F shows the growth biofilm after 3 months following a saline wash step. Cultivation and washing techniques were performed as described in Materials and Methods.
To check the biofilm still consisted of *P. acnes* and not a contaminant, saline washes were conducted weekly and non-adhering cells were spread onto reinforced clostridial and heated-blood agar and incubated anaerobic and aerobic, respectively. The former only returned with colonies with the morphology expected of *P. acnes*, while the latter did not show any growth.

After 4 weeks on incubation, the biofilms were washed and stained with commercial viability assay kit, LIVE/DEAD® BacLight™ Bacterial Viability kit (Invitrogen) and then examined using confocal laser scanning microscopy. The kit contains two DNA-chelating fluorescent dyes. The first, SYTO-9, is membrane permeable and can therefore stain the DNA of viable cells as well as those with perturbed membrane structures. The second dye, propidium iodide can also enter the cell however is readily pumped out using the proton gradient of the cell membrane. Detection of propidium iodide by confocal microscopy indicates damaged cell membrane and proton gradient is lost and the cells are regarded to be non-viable. As shown in Figure 2.4, there is a mixed population of viable and dead cells but the majority appear viable. The same sample was desiccated and visualised using scanning electron microscopy. The specimen was placed on a stage, the chamber was chilled to -20°C and pressure lowered to 50 Pa. This low temperature and low pressure condition is similar to environmental scanning electron microscopy, where sample can be visualised without heavy metal coating. Figure 2.5 shows the comparison between the surface structure of the nitrocellulose substratum and the biofilm. The region of the image with higher cell density is shown by the increase in brightness. From the 1000 x magnification, the biofilm culture showed a combination of void and channel structures similar to what has been previously reported for a *Pseudomonas* biofilm. The structure of the biofilm often reflects on the nutrient availability; commonly biofilms possess a hollow or sponge-like core with channels leading to the surface allowing gaseous exchange and transportation of liquids throughout the depth of the culture (Stoodley *et al.*, 1994, de Beer *et al.*, 1994). At magnification of 5000 x and 10,000 x, these structures can be seen more clearly and formed from tightly ordered short rod-shaped cells.
Figure 2.4 Confocal laser scanning microscopy of *P. acnes* biofilm. Biofilm that was been grown for 4 week was stained with LIVE/DEAD® BacLight™ Bacterial Viability kit (Invitrogen) to assess *P. acnes* viability using confocal laser scanning microscopy. The image was captured every 10 μm from the base of the biofilm until 500 μm was reached. All panels show a top-down view of the stained biofilm, with the captured images collated. Panel A and B show viable cells stained green with SYTO-9 and dead cells stained red with propidium iodide, respectively. Panel C shows the overlay of the two dyes.
See next page for figure legend
Figure 2.5 Scanning electron microscopy of the nitrocellulose substratum and *P. acnes* biofilm.
The surface of biofilm was visualised using low-temperature, low-pressure scanning electron microscopy. Panel A and B show the surface of a new nitrocellulose disk and a nitrocellulose disk on which a biofilm has been cultivated for 4 weeks, respectively. Each vertical panel shows the specimen with increasing magnification (100 x, 1000 x, 5000 x, and 10,000 x). The labelling (white text) at the bottom left of each image indicates from left to right the associated voltage, working distance, signal name (from camera) and atmospheric pressure.

2.3.4 Quality and yield of RNA isolation

To isolate RNA of suitable quantity and quality for transcriptome analysis, a commercial nucleic acid-extraction kit, a protoplasting approach and a published protocol were tested. In all cases, prior to harvesting cells, a mixture of 5% phenol in ethanol was added to cultures to a final concentration of 0.625% phenol to quench cellular metabolism (Lin-Chao & Cohen, 1991). The different protocols utilised the same batch of cells, harvested around 1.0 OD<sub>600</sub> from TYG broth. The commercial kit was RiboPure™-Bacteria (Ambion). The associated extraction method involved homogenising cells suspended in RNA<sub>WIZ</sub>, a phenol containing lysis buffer, with Zirconia beads. Chloroform was added to the lysate and the aqueous phase withdrawn. Nucleic acid was then isolated by precipitating with ethanol, and captured by glass-fibre filtration. Precipitated nucleic acid on the glass fibre was washed with ‘Wash Solutions’ provided by manufacturer and eluted with pre-heated ‘Elution Solution’. Extracted nucleic acid was further treated with DNase I (Ambion) to obtain only the RNA. RNA extraction using this kit produced good quality RNA, as judged by comparison with a sample of *E. coli* total RNA (Figure 2.6). However, the percentage RNA recovered from cells was low: only 5.6 µg of a theoretical amount of 50 µg was isolated from 1.0 OD<sub>600</sub> units of cells (Neidhardt et al., 1987). In addition, the glass-fibre filter that is central to this kit does not capture small RNAs smaller than 200 nt (Figure 2.6), which are now known to have a major role in regulation gene expression in bacteria (Wassarman, 2002).

Also attempted was the extraction of RNA from protoplasts generated by incubating washed cells in a mixture of lysozyme and mutanolysin (Calandra & Cole, 1980). These enzymes degrade
peptidoglycan, more specifically the hydrolysis of the β 1-4 linkage between N-acetylmuramic acid and N-acetyl-glucosamine. The protoplasted cells were then lysed using silica beads (Lysing Matrix B) and Homogeniser Fastprep-24, and then total nucleic acid was extracted using phenol-chloroform and precipitated. The RNA produced by this method was degraded: instead of tight bands corresponding to 23S and 16S rRNA, a smear of smaller RNA fragments was detected (Figure 2.6). The source of the RNA degrading activity was not investigated, as experiments conducted in parallel indicated that a protocol developed for Streptomyces (Kieser et al., 2000, Van Dessel et al., 2004) yielded P. acnes total RNA of sufficient quality and quantity (Figure 2.6). Bands corresponding to 5S rRNA or tRNA were detected in addition to those for 16S and 23S rRNA. Moreover, the yield obtained using this protocol was 3 fold higher than that obtained using the RiboPure-Bacteria kit.

After potassium downshift and the biofilm culture, cells were shown to be still viable albeit with retarded growth. To determine if the integrity of P. acnes RNA under the established culture condition was still suitable for subsequent molecular analysis, RNA was isolated from culture under these conditions and analysed by gel electrophoresis: all were of good quality as judged by the presence of tight bands corresponding to the rRNAs (Figure 2.7).
Figure 2.6 Analysis of total nucleic acid isolated using different techniques by gel electrophoresis. *P. acnes* total nucleic acid was isolated using different techniques. The samples were loaded and analysed by gel electrophoresis; 1.2% (w/v) agarose. All protocols were carried out as described in Materials and Methods. Lane 1 contains purified *E. coli* RNA (a gift from Dr. Louise Kime), while the rest contain *P. acnes* RNA. Lane 2 contains RNA isolated using the RiboPure™-Bacteria (Ambion); lane 3 and 4 contain total nucleic acid isolated using a protocol that incorporated mutanolysin and lysozyme, respectively, and lane 5 contains total nucleic acid isolated using Kirby mix (*Streptomyces* protocol).
2.3.5 Confirmation of kdp operon induction by RT-PCR

Having identified conditions that produce reproducible P. acnes growth and a method for isolating high quality total RNA, the next step was to confirm that the potassium downshift induced the expression of the kdp operon, which as described in the introduction to this chapter encodes a potassium uptake system and an associated two component regulatory system (Ballal et al., 2007). This was approached using RT-PCR analysis (Ballal & Apte, 2005). Primers were designed against PPA0116, a homologue of kdpB (Figure 2.8) and PPA0010 (gyrA), which encodes DNA gyrase subunit A. The latter is considered a housekeeping gene and has been used previously by others as an internal control (Eleaume & Jabbouri, 2004). RNA isolated from duplicate cultures of P. acnes with and without potassium downshift were analysed. The
downshift was mediated as outlined above by growing a culture to mid-exponential phase, harvesting the cells, splitting into two equal halves and washing and reincubating one half in standard HSM and the other in HSM to which potassium had not been added. After 1 h of reincubation, cell metabolism was quenched by adding phenol and total RNA isolated.

As shown in Figure 2.8, the level of the \textit{kdpB} transcript is clearly higher following the potassium downshift, whilst the level of the \textit{gyrA} transcript appears unchanged. Densitometric analysis of the gel revealed that the \textit{kdpA} amplicon was 220-fold higher following potassium downshift (values normalised to \textit{gyrA} amplicon). This represents a minimal fold change as the abundance of the amplicons corresponding to post-downshift samples probably plateaued prior to termination of the PCRs. Both the \textit{kdpB} and \textit{gyrA} amplicons migrated as expected for their predicted size (179 and 85 bp, respectively).
Figure 2.8 Analysis of RT-PCR product from RNA isolated with and without potassium downshift. Complementary DNA was synthesised from equal amount of *P. acnes* RNA with and without downshift, and used as template for PCR amplification of segments of the target genes PPA0010 (*gyrA*) and PPA0116 (*kdpB*) and analysed by electrophoresis using a 2.0% [w/v] agarose gel. Panel A shows comparison between the gene organisation of *kdp* operon in *E. coli* and *P. acnes*. The direction of translation is indicated by the arrow. Forward and reverse primer binding site for the PPA0116 amplicon is boxed in blue and red, respectively. Panel B shows the result of electrophoresis. The different templates used in each lane of the PCR reactions are as indicated; lane M shows a 100-bp DNA ladder (Fermentas), lane 1 and 2 show product of PCR using genomic DNA and no template, respectively. Lane 3 and 4 show the product of PCR using cDNA synthesised from RNA without and with downshift, respectively. Lane 5 and 6 are as lane 3 and 4 but of a biological replicate RNA samples.
2.4 Discussion

The work described in this chapter was successful in growing *P. acnes* reproducibly in Holland synthetic medium (HSM) via batch culture (Figure 2.1), identifying a method for isolating high-quality RNA in reasonable yields (Figure 2.7), and producing a predictable genetic response by means of a potassium downshift (Figure 2.8). This work provides a platform for investigating the response at the level of transcription of *P. acnes* to stresses encountered on the environment of the skin. Initially, it was thought that continuous culture would form the basis of physiological comparisons. Indeed, prior work has described the effects of glucose concentration, oxygen, temperature and pH in *P. acnes* in continuous culture in term of its biomass, maximum specific rate of growth, and extracellular enzyme production associated with the production of substance that may initiate inflammation (lipase, hyaluronate lyase and acid phosphatase) (Cove et al., 1983, Greenman et al., 1981, Greenman et al., 1983). However, when a condition results in slow growth, continuous culture is time consuming and can be particularly cumbersome should, for example, contamination occur. These drawbacks have to be outweighed by the ability to monitor and control multiple growth parameters. In the case of producing transcriptome data that can be mined to uncover mechanisms controlling cellular responses to physiological and environmental changes, most measurements have been made using batch cultures. An example of a transcriptome database dominated by results from cells grown in batch culture is the *E. coli* K12 database, which contains the results of over 3000 microarray experiments and data from 85 publication and is curated by University of Stanford-University of Princeton (Gollub et al., 2003). Mining of *E. coli* transcriptome data from multiple experiments has identified network modules (sets of genes that are co-expressed under some, but not necessarily all, conditions) as well as new regulatory factors both cis and trans acting (De Keersmaecker et al., 2006, Michoel et al., 2009). Rather than to simply collect and analyses further samples from cultures in which growth was affected by altering other external parameters, such as pH and temperature, it was decided that the next priority would be to assess the extent to which the apparent reproducibility of growth in batch cultures is reflected
at the level of the transcriptome and to develop a pipeline of tools to analyse *P. acnes* gene expression and regulation.

Although *P. acnes* is a unicellular organism, over the course of this study cells were found to clump, particularly when growing slowly. This provided a rational for the study of biofilm production. Previous work already indicated the ability of *P. acnes* to stick to biomaterial (silicon, steel titanium and plastic) by production of an exopolymer similar to the polysaccharide intercellular adhesion of *S. aureus* (Bayston et al., 2007, Ramage et al., 2003). By growing *P. acnes* on nitrocellulose disk, it was possible to produce a lawn of cells that were adhered to the surface of substratum. The viability of the biofilm culture was confirmed by fluorescence staining and confocal microscopy. High magnification under SEM revealed sheet- and channel-like structure formed from *P. acnes* cells, confirming the formation of a biofilm (Figures 2.3 & 2.4). Biofilms are mostly composed of carbohydrates, the UDP-N-acetylglucosamine-2-epimerase and glycosyl transferase has been hypothesised to play a role in synthesis of the glycocalyx polymer that constitutes *P. acnes* biofilm (Burkhart & Burkhart, 2003). Indeed from the annotated *P. acnes* genome, it was found to posses UDP-N-acetyl-D-mannosaminuronate dehydrogenase, UDP-N-acetylglucosamine2-epimerase, mannose-1-phosphate guanylyltransferase, ExoA (succinoglycan biosynthesis protein), and various glycosyl transferases found in at least 3 gene clusters PPA125-134, PPA145-150, PPA1692-1700. Quorum-sensing is major part of regulating and initiating biofilm formation by synthesis of an autoinducer-2 signal molecule. This was shown by mutational studies of known biofilm forming organisms such as *Streptococcus mutans*, and *Staphylococcus epidermidis* (Merritt et al., 2003, Xavier & Bassler, 2003, Xu et al., 2006). Coeyne performed a comparative genome analysis of the *P. acnes* genome to that of *Vibrio harveyi* from which the quorum-sensing system, LuxS, was first characterised (Coeny et al., 2007). It was found that PPA0450 is a homolog to LuxS, however genes involved in the signal transduction pathway of LuxS were not found in *P. acnes*. Differential gene expression profiling of *S. aureus* biofilm and planktonic cultures using microarrays revealed gene clusters involved in cell wall synthesis, polysaccharide intercellular adhesins, and stress response proteins were significantly up-regulated. Over 200 hypothetical
genes with unknown function were also shown to be up-regulated (Resch et al., 2005). Similar differential analyses were carried out in P. aeruginosa as the model biofilm forming organism with the aid of RNA-sequencing (Dotsch et al., 2012). The work was able to develop multiple expression profiles of planktonic and biofilm cultures that correlated well to previously published work. In addition, over 600 putative transcriptional start sites and a 31 small RNA were reported to be expressed under biofilm conditions, which could not be identified by microarray analysis. It would therefore be interesting to compare the profile of gene expression in P. acnes biofilms not only with that of cells grown in liquid culture, but the biofilms of other bacterial species.

Much of the work described herein also provides a platform for studying gene expression at the level of the proteome. Recent advances in mass spectrometry have dramatically increased the sensitivity, coverage and throughput of this approach (Cox & Mann, 2011, Wright et al., 2012). Ultimately, it would be interesting to compare the results of analysing gene expression at the level of the transcriptome and proteome. Such comparisons have been done for other bacteria, including Streptomyces coelicolor, which like P. acnes is an actinomycete. Jayapal et al. have shown the GroEL stress protein showed a discordant pattern in mRNA and protein expression level; the transcript level of groEL decreased with increased incubation time whilst the protein remains at a similar level of abundance throughout, suggesting post-translational modification of the protein allowing adaptation to different phases of growth (Jayapal et al., 2008). Thomas et al. investigated metabolic switches and adaptation of the bacterium to the deletion of phoP in S. coelicolor. They observed the change in carbon source from glucose to glutamate when phosphate was depleted; the level of enzymes involved in gluconeogenesis is high compared to the wild-type. The absence of the PhoP-regulated protein and its knock-on effect on oxidative phosphorylation resulted in an imbalance on the ratio of NAD/NADH. This led to the hypothesis that S. coelicolor utilises gluconeogenesis as a way to compensate for the imbalance in the ratio of NAD/NADH (Thomas et al., 2012). Although proteomics adds another dimension to the analysis of gene expression, regulation at the level of translation, can still be detected at the level of transcription. For bacteria, it is well established that there is interplay between
translation and mRNA degradation (Yarchuk et al., 1992). Changes that reduce the coverage of mRNA by ribosomes increase susceptibility to attack by ribonucleases (Carpousis et al., 2009a). This is illustrated most recently by the finding that mRNAs bound by antisense RNAs that block translation are degraded more rapidly (Storz et al., 2004, Masse et al., 2003). This increased degradation although a secondary effect, which might reinforce effects on translation, is reflected at the level of transcript abundance. Thus, it cannot be assumed that differential gene expression observed at the level of the transcriptome reflects changes at the level of transcription initiation. For further details of bacterial gene regulation post-transcriptional initiation, readers are directed to several excellent reviews (Nogueira & Springer, 2000, Gold, 1988, Arraiano et al., 2010, Timmermans & Van Melderen, 2010).
Chapter 3

3 Analysis of the global transcriptional responses of *P. acnes* to potassium-downshift

3.1 Introduction

The field of genetics up until the mid 1990's was largely focussed on identifying genes associated with specific phenotypes, such as the ability to undertake particular biochemical transformations or to facilitate transitions in the cell-cycle or developmental pathways. With the advent of genome sequencing and the cataloguing of thousands of gene families, the emphasis shifted towards discovering the function of previously uncharacterised genes and obtaining a more holistic view of how different genes interact to mediate cellular and developmental processes (Wang *et al.*, 2009a). One approach was to describe the expression patterns of entire genomes, as genes that interact functionally tend to be co-expressed. This approach received a major boost with the description of DNA microarrays by the laboratory of Pat Brown in the ‘Genome Issue’ of Science (Schena *et al.*, 1995). This technology, which allowed the expression of thousands of genes at the RNA level to be measured in parallel, far surpassed what had been achievable using nuclease mapping (Berk & Sharp, 1977), primer extensions (Shelness & Williams, 1985), dot blots (Kafatos *et al.*, 1979) and macroarrays, its immediate predecessor (Wada *et al.*, 1999). Over the last decade and a half, microarrays have been used extensively to study organisms ranging from bacteria to humans (Bier & Kleinjung, 2001, Goldsmith & Dhanasekaran, 2004).

Almost all of the microarrays that are used today are supplied by commercial manufacturers and contain probes that are synthesised *in situ*. Over the years, the density of probes has increased to allow transcriptional measurement to be extended to all regions of the genome, not just annotated genes. Indeed, arrays can now cover 3 Mbp with probes every 10 or so bp
on each strand. These high-density ‘tiling’ arrays enabled the discovery and study of untranslated regions, non-protein-coding RNAs, alternative transcriptional units and RNA processing. Today, single-nucleotide resolution is achievable using global RNA-sequencing (gRNA-seq), also called whole transcriptome shotgun sequencing (Mamanova & Turner, 2011). Currently, Illumina Solexa is a popular platform for the actual sequencing steps of RNA-seq (Marioni et al., 2008). However, as higher-throughput technologies are developed, these will replace Illumina Solexa. Indeed, there are reports that single molecule direct RNA sequencing (DRSTM) technology is currently being developed by Helicos (Pushkarev et al., 2009).

The adoption of RNA-seq for transcriptomic studies will almost certainly lead to an unparallel expansion in the number of gene expression profiles, which for many model species, such as the bacterium *Escherichia coli*, are being collated in single compendia. This in turn should increase the power of computation approaches to identify the networks of gene interactions that mediate complex functions within the operational context of the whole cell. Moreover, single-nucleotide resolution should ease the identification of cis sequences shared by genes that are co-expressed. Having established growth conditions for *P. aenes*, this chapter first describes the use of microarrays to determine the extent to which reproducible growth is reflected at the transcriptome level and to chart the response of this organism to a model stimulus, potassium salt downshift. It then goes on to describe the adoption of a global RNA-seq approach and its benefits over microarrays.
3.2 Materials and Methods

3.2.1 Gene expression microarray

Samples of total RNA were isolated and purified from duplicate cultures cultivated in HSM with and without potassium downshift (2 x cultures, 2 x conditions), as described in Chapter 2. The samples were then sent to Roche NimbleGen (Iceland), where they were analysed using their single-channel system and quadruplex chips; each of the arrays containing 72k probes; 16 probes per target gene and two probe sets per array. The probes were 60mer oligonucleotides. The array design name was T1267747_60mer for *Propionibacterium acnes* strain KPA171202 with probes designed using genomic data as detailed in NC_006085. Probe intensities were collected and normalised using the algorithm Robust Multi-array Average (RMA) (Irizarry et al., 2003) also performed by Roche NimbleGen. For each of the four RNA samples, the normalised average probe intensity per gene for both probe sets was provided as a tabular delimited text file. Each of these datasets was then compared using M-A (ratio-intensity) scatterplots. For each A value, we calculated the average (μ) and standard deviation (σ) of M in a moving window of 100 pairs that were sorted in ascending order of A. Upper and lower envelopes were defined by the equation: \( \mu \pm 3\sigma \), and positions outside the envelope recorded, as described previously (Hovatta et al., 2005, Marincs et al., 2006). Details of specific comparisons are provided in the Results section.

3.2.2 Rank Product analysis

The statistical significance of the normalised probe intensity data were calculated using RankProdlt (http://strep-microarray.sbs.surrey.ac.uk/RankProducts/). This is an online tool that utilises the Rank Product/Rank Sum algorithm (Breitling et al., 2004), which can identify differently expressed genes from two or more replicates. After the output was generated, lists of differentially expressed genes were obtained by sorting using the column ‘probability of false positive value (pfp)’ with the cut-off value of pfp < 0.15.
3.2.3 Global RNA sequencing

The same RNA samples sent for microarray analysis were enriched for mRNA using MICROBExpress™-Bacteria oligocapture magnetic beads, as described by the manufacturer (Ambion). Global transcriptome sequencing was performed by Dr. Lira Mamanova (Wellcome Trust, Sanger Institute, Cambridge, UK) using enriched mRNA and a published methodology (Mamanova et al., 2010a). Sequencing was done using an Illumina Solexa platform. RNA sequences from the global analysis were processed in-house using Galaxy (Goecks et al., 2010) and mapped to the genome using Bowtie 2.0 (Langmead & Salzberg, 2012) with custom parameter (-y -a -best -strata).
3.3 Results

3.3.1 Analysis of the transcriptome for differentially expressed gene using gene expression array and Rank Product algorithm

To obtain a global view of the *P. acnes* transcriptome in response to the potassium downshift (Chapter 2), total RNA samples were sent to Roche NimbleGen (Iceland), where they were analysed using their quadruplex chips and a two-channel system. Pairwise comparisons were then performed using M-A (ratio-intensity) scatterplots (Figure 3.1). Comparison of the results for the duplicate cultures without or with the potassium downshift (Figure 3.1, panels A and B, respectively) revealed that the vast majority of the M (ratio) values were close to 0, indicating that there was no major difference in the global gene expression of the two cultures. This is consistent with the high level of reproducibility which was obtained in cell culturing (Figure 2.2, Chapter 2). Similar results, with one important exception, were also obtained when gene expression was compared between cultures with and without potassium downshift (Figure 3.1, panels C and D, respectively). The exception being that a small number of genes had M values well above the general scatter of points, i.e. had increased gene expression following potassium downshift. An initial inspection revealed that this group included genes of the *kdp* operon, as expected. Further analyses of changes in gene expression as a result of the downshift are described below. It should be noted that because each array contained a duplicate set of probes, each gene is represented by two pairs of M-A values. Our overall interpretation of the downshift data is that sampling after 1 hour allowed detection of a specific response, without sufficient time having elapsed for consequences on growth to be manifested. The re-incubation period allowed sufficient time for *P. acnes* to respond to the potassium down-shift but not be affected by the removal of phosphate from the medium.

Genes with altered expression as a result of the downshift were identified using two published approaches. The first was based on analysing the M-A scatterplots shown above, taking into account the scatter of points obtained when comparing biological replicates (Figure 3.1, panel A and B). The vast majority of the points in each comparison (with or without downshift) could
be contained with envelopes described by the equation $\mu \pm 3\sigma$, where average ($\mu$) and standard deviation ($\sigma$) of the M values are within a window of A values that slides from the lowest to highest value (i.e. moves left to right along the y-axis). No genes were found to be outside these 'noise' envelopes in both comparisons of the biological replicates (data not shown). In contrast, 32 genes were found to be outside the envelopes in each of the duplicate comparisons to determine the effects of the downshift (Figure 3.1, panels C and D). These genes and the corresponding fold changes are listed in Table 3.1. In addition, we analysed the microarray data using an online version of Rank Product algorithm (Laing & Smith, 2010), which detects differentially regulated genes in replicated microarray experiments. The filtering criteria for rank product analysis are based on the false positive discovery rate rather than a p-value of the comparison between two conditions. This is because with an increase in sample population there is also an increase in false discovery rate (Breitling & Herzyk, 2005). Many of the genes identified with altered expression by the analysis of the M-A scatterplots were also identified by Rank Product analysis (Table 3.1). These included all the genes of the kdp operon. Overall the analysis of the M-A scatterplots appears to have been more sensitive. In two cases, it identified all of the genes in a cluster with related function, while Rank Product did not (see PPA1287-90 and PPA1758-60).
Figure 3.1 M-A scatterplot of the gene expression values from microarray analysis. In panel A, the M values equal $\log_2$ (without-downshift sample 2/ without -downshift sample 1), and A equals $(\log_2$ without -downshift sample 2 + $\log_2$ without -downshift sample 1)/2. Panel B as A, but substituted values from with-downshift. In panel C, M values equals $\log_2$ (with- downshift/without-downshift), and A equals $(\log_2$ with-downshift + $\log_2$ without-downshift)/2. Panel D as C, but biological duplicate. The circled region in panel C and D highlight genes significantly upregulated from the effect of potassium down-shift. The red and blue points represent the upper and lower boundaries of the ‘noise’ envelopes. For each A value, the average ($\mu$) and standard deviation ($\sigma$) were calculated from a sliding window of the 100 corresponding A values. The upper and lower envelopes were defined by the equation: $\mu \pm 3\sigma$. 
### Table 3.1 Genes showing significant change in expression with and without potassium downshift.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Function</th>
<th>Increase (+) / Decrease (-)</th>
<th>Fold change in expression</th>
<th>Probability of false positive &lt; 0.15</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPA0114</td>
<td>hypothetical protein</td>
<td>+</td>
<td>201.4</td>
<td>+</td>
</tr>
<tr>
<td>PPA0115</td>
<td>potassium-transporting ATPase subunit A</td>
<td>+</td>
<td>242.3</td>
<td>+</td>
</tr>
<tr>
<td>PPA0116</td>
<td>potassium-transporting ATPase B chain</td>
<td>+</td>
<td>224</td>
<td>+</td>
</tr>
<tr>
<td>PPA0117</td>
<td>potassium-transporting ATPase C chain</td>
<td>+</td>
<td>69.1</td>
<td>+</td>
</tr>
<tr>
<td>PPA0118</td>
<td>two-component sensor, KdpD</td>
<td>+</td>
<td>19.5</td>
<td>+</td>
</tr>
<tr>
<td>PPA0119</td>
<td>two-component response regulator, KdpE</td>
<td>+</td>
<td>33.9</td>
<td>+</td>
</tr>
<tr>
<td>PPA0120</td>
<td>hypothetical protein</td>
<td>+</td>
<td>5.9</td>
<td>+</td>
</tr>
<tr>
<td>PPA0476</td>
<td>glucosamine-6-phosphate isomerase</td>
<td>+</td>
<td>4.6</td>
<td>-</td>
</tr>
<tr>
<td>PPA0667</td>
<td>hypothetical protein</td>
<td>+</td>
<td>2.5</td>
<td>-</td>
</tr>
<tr>
<td>PPA1091</td>
<td>hypothetical membrane associated protein</td>
<td>+</td>
<td>3.1</td>
<td>-</td>
</tr>
<tr>
<td>PPA1092</td>
<td>protein with SuSA/YciO/YrdC domain</td>
<td>+</td>
<td>2.8</td>
<td>-</td>
</tr>
<tr>
<td>PPA1093</td>
<td>S'-methylthioladenosine/S-adenosylhomocysteine nuclosidase</td>
<td>+</td>
<td>2.4</td>
<td>-</td>
</tr>
<tr>
<td>PPA1105</td>
<td>phosphoglucomutase</td>
<td>+</td>
<td>2.1</td>
<td>+</td>
</tr>
<tr>
<td>PPA1224</td>
<td>phosphodiesterase</td>
<td>+</td>
<td>2.9</td>
<td>-</td>
</tr>
<tr>
<td>PPA1287</td>
<td>non-ribosomal peptide synthetase</td>
<td>+</td>
<td>2.9</td>
<td>-</td>
</tr>
<tr>
<td>PPA1288</td>
<td>surfactin synthetase subunit 1</td>
<td>+</td>
<td>3.4</td>
<td>-</td>
</tr>
<tr>
<td>PPA1289</td>
<td>cysteine synthase/ornithine cyclodeaminase</td>
<td>+</td>
<td>3.6</td>
<td>+</td>
</tr>
<tr>
<td>PPA1290</td>
<td>cystathionine beta-synthase</td>
<td>+</td>
<td>4.1</td>
<td>+</td>
</tr>
<tr>
<td>PPA1676</td>
<td>ferrous iron transport protein A</td>
<td>+</td>
<td>3.8</td>
<td>-</td>
</tr>
<tr>
<td>PPA1677</td>
<td>ferrous iron transport protein B</td>
<td>+</td>
<td>3.8</td>
<td>-</td>
</tr>
<tr>
<td>PPA1758</td>
<td>outer membrane lipoprotein</td>
<td>+</td>
<td>5.9</td>
<td>+</td>
</tr>
<tr>
<td>PPA1759</td>
<td>ABC transporter ATP-binding protein</td>
<td>+</td>
<td>5.0</td>
<td>+</td>
</tr>
<tr>
<td>PPA1760</td>
<td>ABC transporter associated permease</td>
<td>+</td>
<td>3.8</td>
<td>-</td>
</tr>
<tr>
<td>PPA2286</td>
<td>phosphoglucomutase/phosphomannomutase</td>
<td>+</td>
<td>2.9</td>
<td>-</td>
</tr>
<tr>
<td>PPA2268</td>
<td>alanine dehydrogenase</td>
<td>+</td>
<td>1.6</td>
<td>+</td>
</tr>
<tr>
<td>PPA0012</td>
<td>L-lactate dehydrogenase</td>
<td>-</td>
<td>2.3</td>
<td>-</td>
</tr>
<tr>
<td>PPA0557</td>
<td>sodium-and chloride-dependent transporter</td>
<td>-</td>
<td>3.8</td>
<td>-</td>
</tr>
<tr>
<td>PPA0964</td>
<td>transcriptional regulator</td>
<td>-</td>
<td>2.9</td>
<td>-</td>
</tr>
<tr>
<td>PPA1476</td>
<td>glycine betaine transport system permease protein</td>
<td>-</td>
<td>2.1</td>
<td>-</td>
</tr>
<tr>
<td>PPA1807</td>
<td>hypothetical protein</td>
<td>-</td>
<td>4.2</td>
<td>-</td>
</tr>
<tr>
<td>PPA2152</td>
<td>putative peptide transport system secreted</td>
<td>-</td>
<td>3.0</td>
<td>-</td>
</tr>
<tr>
<td>PPA2175</td>
<td>peptide-binding protein</td>
<td>-</td>
<td>4.1</td>
<td>+</td>
</tr>
</tbody>
</table>
Analysis of the microarray data revealed that genes with the largest increase in expression were the structural components of the kdp operon PPA0115, PPA0116 and PPA0117 (242.3, 224 and 69.1 fold respectively). The genes of the sensor kinase (PPA0118) and response regulator (PPA0119) were also expressed at a higher level (19.5 and 33.9 fold, respectively), but not as high as those of the structural components of the potassium-uptake system. The fold changes in the structural genes correlate well with the result of RT-PCR analysis described in Chapter 2, but lower than that reported for E. coli and Salmonella spp. (Frymier et al., 1997, Hamann et al., 2008). In the absence of potassium downshift, the level of transcription of kdpDE is higher than the upstream genes of the operon. This would explain why, following potassium downshift, the fold increase associated with kdpDE is not as high.

Unexpectedly, two adjacent genes were also found to be up regulated PPA0114 (201.4 fold) and PPA0120 (5.9 fold), both of which are annotated as encoding hypothetical protein. From the data alone, this suggests that PPA0114 and PPA0120 are part of the kdp operon in P. acnes, the remainder of which is highly conserved amongst bacterial species. Ancillary proteins have been identified in other bacteria, for example, the first gene in the kdp operon of E. coli contains a small protein (KdpF) that stabilise the KdpATPase-pump (Gassel et al., 1999). PPA0114 and PPA0120 are not however homologues of this E. coli protein.

Genes encode protein involved in production of surfactants were also found to be up regulated. These genes encode non-ribosomal peptide synthetase (NRPS; PPA1287), cysteine synthase/orinithine cyclodeaminase (PPA1289) and cystathionine beta-synthase (PPA1290). The microarray data of the putative NRPS gene cluster (PPA1286 to PPA1291) showed a 3 fold increase in expression on average after potassium downshift. Studies of non-ribosomal peptides have shown that they can have high affinity for iron and fulfil a role as iron chelators in bacteria (Challis & Naismith, 2004). Other genes with increased expression were also linked to iron homeostasis; the ferrous iron transport proteins A and B (PPA1676 and PPA1677, respectively) showed on average a 4 fold increase. The simplest explanation is that K2HPO4 and KH2PO4 contained trace amount of contaminating iron, which when removed necessitates
increased production of an associated chelator and transporter. Interestingly, studies have reported that *Pseudomonas* spp. and *S. aureus* form biofilms under iron-limiting conditions to increase the acquisition of iron (Banin *et al.*, 2005, Lin *et al.*, 2012). This may explain the earlier observation (Chapter 2) that cells tended to clump when cultivated in the absence of added phosphate, which was likely a source of iron.

The sodium- and chloride-dependent transporter (PPA0557) and glycine betaine transport system permease protein (PPA1476) showed a 3.7 fold and 2.2 fold decrease in expression. Glycine betaine is an important osmoprotectant (Robert 2000). It is likely that the observed changes reflect the reduced salt concentration of the media as a result of the potassium downshift, which would render the cells hypotonic. Therefore, *P. acnes* would no longer need to accumulate glycine betaine to the same level. Reasons for the altered expression of the remaining genes remain obscure.

### 3.3.2 Analysis of *P. acnes* transcriptome using global RNA-sequencing

To obtain single nucleotide-resolution transcription maps of *P. acnes* before and after potassium downshifts, the RNA samples described above were also analysed using a new RNA-sequencing approach that does not require a PCR amplification step (Mamanova *et al.*, 2010a). This was performed at the Wellcome Trust Sanger Institute by Dr. Lira Mamanova. Samples enriched for mRNA were fragmented by metal ion hydrolysis, dephosphorylated and then rephosphorylated to produce 5'-monophosphated ends to which an adaptor could be ligated; a sequencing adaptor was also ligated to the 3' end of RNA. The adaptors contained both DNA and RNA sequences; the former allows fragments to be attached to the flow-cell surface via hybridisation, while the latter allows the binding of an oligonucleotide that primes the synthesis of cDNA, which is then sequenced directly without PCR amplification. The sequences, which were generated using an Illumina Solexa platform, were then processed in Leeds using Galaxy (Goecks *et al.*, 2010), and mapped to the *P. acnes* genome using Bowtie 2.0 (Langmead & Salzberg, 2012). The number of times each position in the genome was read by gRNA-seq was
then calculated and the results viewed using the UCSC Microbial Genome Browser (Quinlan & Hall, 2010, Schneider et al., 2006). This data provided an independent measurement of the response of P. acnes to the potassium downshift and extended the analysis to beyond annotated protein-coding genes. The later will be described in more detail in the next chapter.

3.3.2.1 Comparison of gene expression values between global RNA-sequencing and microarray.

To determine to what extent the gRNA-seq data was comparable with the microarray results, the density of sequence reads (total number corrected for length) within the coding region was calculated for each gene. These density values were then plotted against the expression values determined using microarrays for the samples corresponding to before and after the potassium downshift. A positive correlation between the gRNA-seq and microarray data was obtained; however, there was a significant amount of scattering (Figure 3.2). Extreme outliers include PPA0971, PPA1877, and PPA2388, which are all annotated as hypothetical proteins. Expression of PPA0971 was 100 fold higher by gRNA-seq analysis. However, inspection of the sequencing data for PPA0971 using the UCSC Microbial Genome Browser revealed that the 3’ coding region of PPA0971 overlaps that of a tRNA, which showed high level of expression (Figure 3.3, panel A). The expression of the tRNA gene did not complicate the analysis of the microarray data, as probes were not designed for the overlapping region. The above indicates that for a maximum value to be extracted from the gRNA-seq data it has to be viewed in the context of gene annotation.

PPA1877 and PPA2388 are representative of genes found to have higher an apparent expression by microarray analysis. Visualisation of these and other genes using the Genome Browser found that expression was dominated by sequencing reads on the opposite, non-coding strand (Figure 3.3, panel B and C). The apparent higher expression by microarray analysis reflects the fact that the NimbleGen service did not provide a strand-specific analysis. When the average gene expressions for gRNA-seq were calculated from both coding and non-
coding strand, the two datasets showed tighter correlation and there was no gene found that showed higher expression by microarray analysis (Figure 3.2, panel C and D). PPA0971 is an outlier for the explanation provided above; the other data point PPA2027, which is to the right of PPA0971, also showed a significant increase in expression from the gRNA-seq data. For PPA2027, a tRNA overlaps with the coding region of PPA2027 on the opposite strand (data not shown) therefore increasing the average reads when both strands were used for calculation.
Figure 3.2. Scatter plot of average expression of protein coding genes from microarray and global RNA-sequencing. Start and end positions of the protein coding regions for *P. acnes* were obtained from NCBI database. The Average expression per gene was obtained by dividing the sum of the sequencing values, defined by the coding region, by the length of the respective gene. Panel A and B contained gene expression values from RNA isolation from *P. acnes* cultured in HSM without and with potassium downshift, respectively. The average gene expression values from gRNA-seq were calculated from the coding strand only. Panel C and D as A and B, except the average gene expression values from gRNA-seq were calculated using both coding and non-coding strand. The $R^2$ value line indicates the correlation between the two dataset. The red and blue circles indicate the extreme outliers, where average gene expression values were significantly high in RNA-sequencing and microarray, respectively.
See next page for figure legend
Figure 3.3. Global RNA-sequencing data of genes with significant difference in expression levels compared to microarray. Data obtained from gRNA-seq was processed as described in Materials and Methods and visualized using the UCSC Microbial Genome Browser. Tracks shown are the number of time each nucleotide has been sequenced. The forward and reverse strand are coloured in black and red, respectively. The numbers of sequencing reads are indicated on the left. Numbers at the top indicate the nucleotide position in the chromosome. Non-customised tracks provided show the protein coding, tRNA, rRNA and non-coding RNA region. Panel A, B and C show tracks for PPA0971, PPA1877, and PPA2388, respectively.

3.3.2.2 Analysis of potassium response by global RNA-sequencing

Having confirmed the gene expression profile obtained from gRNA-seq is broadly comparable to microarray, the next step was to analyse further the genes identified as being differentially expressed by the microarray analysis. Analysis of PPA0114 and PPA0120 suggested that transcription of the kdp operon following downshift starts within the former. Consistent with this notion, analysis of the sequence immediately upstream of the leading edge of transcription revealed a possible -10 box promoter and a direct repeat that could be the binding site for KdpE (Figure 3.4, panel A). In light of the gRNA-seq data, PPA0114 does not appear to be part of the kdp operon. In the absence of potassium downshift PPA0120 appears to be transcribed independently of the upstream genes: there is a step increase in transcription immediately 5' to PPA0120. This suggests that this gene could function independently of the potassium uptake systems. The increase in transcription following downshift may result from inefficient termination immediately downstream of kdpE (PPA0119). In the absence of potassium downshift, a leading edge can be detected upstream of kdpD. This likely corresponds to the promoter of a nested transcription unit that is obscured when transcription is induced from the promoter within PPA0114.

Characterisation of the KdpE binding site in E. coli upstream of the kdpABC transcriptional start site, but not upstream of the kdpDE transcriptional start site further indicated an alternative
transcriptional regulation of kdpDE (Sugiura et al., 1993, Sugiura et al., 1992). In the presence of potassium, kdpDE was found to be transcribed as a transcriptional unit independent of kdpABC in E. coli. The transcriptional start site for the putative kdpD transcript in P. acnes correlated well with the work published in E. coli (75 bp upstream of the start codon of KdpD; (Polarek et al., 1992). Inspection of the 3' end nucleotide sequences of PPA0117 (KdpC) revealed a motif (TAAGGT) that resembles a -10 consensus sequence 14 bp upstream of the kdpD transcript (chromosome position 142112). However no motifs that resemble -35 consensus sequence could be found. This suggests kdpDE are regulated by a vegetative promoter, the poor matches to -35 consensus sequence limit the level of expression to a basal level as the cell only requires kdpDE to be expressed during osmotic shock.

The gRNA-seq data also shows a transcriptional unit encompassing the putative NRPS gene cluster (PPA1286 to PPA1291) (panel B). Moreover, the fold increase in expression was similar to that detected using microarrays. In the absence of down-shift, a step increase in transcription was observed at the 5' end of PPA1287. This suggests independent regulation of the PPA1286-87 transcript from the PPA1291 transcriptional unit by a nested promoter. However, from the data alone, the step increase in transcription could correspond to a site of processing as well as transcriptional initiation.
See next page for figure legend
Figure 3.4. Detection of *P. acnes* gene expression by global RNA-sequencing. Data obtained from gRNA-seq was processed as described in Materials and Methods and visualized using the UCSC Microbial Genome Browser. Labelling as Figure 3.3. Panel A and B shows the putative *kdp* and non-ribosomal peptide synthetase cluster, respectively. Red arrows in panel A and B mark positions that suggest a separate transcriptional unit. The top insert of panel A shows the first 100 bp nucleotide of PPA0114. The nucleotide coloured in red labels the putative start of the *kdpABC* transcriptional unit, denoted as +1. The double-ended arrow shows the putative binding site of PPA0119. The top inset shows the direct repeat sequence, putative -10 and -35 regions are underlined. The bottom inset of panel A is a magnified view of PPA0118 and PPA0119 at a different scale.
3.3.2.3 Identification of differentially expressed genes by gRNA-seq

The analysis described above demonstrates the sensitivity of the gRNA-seq approach. To determine if the downshift results in increased transcription of genes not covered by the NimbleGen arrays, the gRNA-seq data was analysed using an M-A scatterplot method (Figure 3.5). The analysis was performed by comparing the sequence reads obtained at each nucleotide position with and without the potassium downshift. Forward and reverse strand were analysed separately (Panels A and B, respectively). Most of the M (ratio) values scattered around 0, and there was a clear group with M values greater than 0 (forward strand, panel A). This is consistent with the results of the microarray analysis; the most induced genes (kdp and NRPS operons) are encoded on the forward strand. Unexpectedly a large group with M values less than 0 for both strands was observed. To analyse the gRNA-seq further, an envelope was defined as described previously, and values above and below were selected for further analysis. The nucleotide positions corresponding to these values were then determined (UCSC Microbial Genome Browser). This analysis revealed that M values less than 0 corresponded to 'stacks' of short sequencing reads in the ‘without-downshift’ library (Figure 3.6). The origin of the stacks is unknown, but may be an artefact of gRNA-sequencing. The stacks were found throughout the genome on both strands (data not shown). The presence of these stacks complicates the screening of the gRNA-seq data for genes with altered expression as a result of potassium downshift. A set of rules to specifically eliminate stacks has not yet been defined. However, manual inspection of all the nucleotides positions with M values above the envelope did not identify any additional genes to those already identified by microarray analysis.
Figure 3.5 M-A scatter plot of global RNA-sequencing reads. Generation of A and M values and the ‘noise’ envelope (red and blue trace) were as described in Figure 3.1, except using the sequencing reads. The average and standard deviation were calculated from a sliding window of the 5000 corresponding A values. Panel A and B show the comparison of the RNA-sequencing reads between RNA samples with and without downshift on the forward and reverse strand, respectively.
Figure 3.6. Identification of position showing differential expression by gRNA-seq in *P. acnes*. Track shows the position identified by M-A scatterplot to be outside the defined noise envelop. Labelling as Figure 3.3. Forward and reverse strand are coloured in black and red, respectively. For each strand, gRNA-seq reads for with and without potassium down-shift are shown in separate tracks. Colours of the annotated genes indicate the cluster of orthologous group (COG). For more detail please see [http://www.ncbi.nlm.nih.gov/COG/](http://www.ncbi.nlm.nih.gov/COG/).
3.4 Discussion

This chapter was successful in establishing that the conditions used to culture *P. acnes* (Chapter 2) are sufficiently reproducible for specific genetic response(s) to be detected readily at the level of the transcriptome. Analysis of the microarray data using M-A scatterplots showed a high degree of similarity in the expression profile between biological duplicates (Figure 3.1). The genetic responses of potassium down-shift were also determined (Table 3.1). As expected, it revealed the increase expression of the Kdp system (PPA0115-PPA0119), which is responsible for sensing and uptake of potassium. Unexpectedly, increased expression was also detected of genes involved in iron homeostasis via the synthesis of non-ribosomal peptide-based iron chelators (PPA1286-PPA1291) and the uptake of iron (PPA1676-PPA1677). It is postulated that the K$_2$HPO$_4$ and KH$_2$PO$_4$ that were removed were the source of trace amount of iron. The contamination of an AnalaR®-grade chemical with iron is precedent. Groups studying the iron regulation use iron-chelators to remove contamination from solutions used to make defined media. As mentioned above, the removal of an iron source may relate to the observation that cells tended to clump when cultivated following downshift, as physical association of cells increases iron acquisition (Banin et al., 2005, Lin et al., 2012). The up-regulation of genes for iron uptake from microarray data suggests physiologically, *P. acnes* is preparing for biofilm formation. This can be investigated further by harvesting the cells for transcriptome analysis after longer exposure to the down-shift. During biofilm formation, cells are able to specialise their metabolism to behave as a community. Scarce nutrients can be concentrated and recycled within the depth of the biofilm. This mode of growth also causes slower metabolism, which reduces the amount of nutrient required (Hall-Stoodley et al., 2004, Stewart & Franklin, 2008).

Global RNA-seq was also used to provide an independent measurement of the transcriptome profile, showing a clear positive correlation with the expression levels obtained from microarray (Figure 3.2, panel A and B). However, the increased resolution revealed important features not detected using microarrays. For example, it showed that PPA0114 is not part of
the kdp operon. Transcription starts within, and not upstream of, this gene (Figure 3.4). It was also revealed that kdpD and kdpE (the sensory components) are transcribed independently of the uptake components prior to the potassium downshift, and suggested that PPA0120, the gene immediately downstream of kdpE, can be transcribed independently of the upstream kdp genes. In addition, the increased resolution allowed the identification of a direct repeat sequence that may be the binding site of KdpE. Further experiments will be required to confirm that this is the case. Viewing of the gRNA-seq data also confirmed that the genes identified by analysing values outside the ‘noise’ of M-A scatterplots did indeed have altered expression (data not shown). Thus, although this method has not been as widely adopted as Rank Product Analysis, it appears more sensitive. Rank Product analysis only detected some of the genes within regulated operons (Table 3.1).

Another major advantage of gRNA-seq is the strand-specificity; the NimbleGen data was only specific for gene loci (Figure 3.2, panel C and D; Figure 3.3). The gRNA-seq data was also as sensitive. NimbleGen have terminated its array service from June 2012, as the market is slowly being replaced by RNA-sequencing. Companies that provide transcriptome profiling are shifting towards RNA-sequencing and the improvement of the technology will make RNA-sequencing even more cost-effective.

A limitation of the gRNA-seq approach (which is shared with microarrays) is that it does not readily identify nested transcription units. For example, the leading of the kdpDE transcription unit was obscured by upstream transcription upon potassium downshift (Figure 3.4). RNA-seq approaches have been developed to identify the 5’ ends of nascent transcripts, differentiating them from those of processing and degradation intermediates. All the published differential RNA-seq approaches utilise a 5’-monophosphate-dependent, 5’ to 3’ exonuclease (Terminator™ 5’ phosphate-dependent exonuclease, TEX), and a pyrophosphatase (tobacco acid pyrophosphatase, TAP) (Sharma et al., 2010). TEX digests transcripts carrying a 5’ monophosphate, leaving 5’ hydroxylated and nascent triphosphorylated transcript untouched (He et al., 2010). Pyrophosphatase decaps nascent transcripts and convert RNA from
triphosphorylated to monophosphorylated form where sequencing adaptors can then be ligated (Levin et al., 2010). Determination of transcriptional start site has shown to add value to existing transcriptional networks as well as the construction of novel ones (Dotsch et al., 2012, Sharma et al., 2010). By understanding the locations of promoters, the transcription factor binding sites and consensus sequences for regulators can be elucidated to improve the understanding of transcriptional network (Salgado et al., 2013, Martin et al., 2010a). Differential RNA-seq analysis of the *P. acnes* samples is described in the next chapter.
Chapter 4

4 Primary and secondary transcriptome analysis of *P. acnes*

4.1 Introduction

The 2.56 Mbp genome of *P. acnes* strain KPA171202 contains 2,333 putative genes of which 87 were annotated as encoding transcriptional regulators (Bruggemann et al., 2004). Prior to the work reported here, no transcriptional start sites (TSSs) had been mapped experimentally. Other key aspects of gene regulation for which no information was available were mRNA turnover, which ensures translation follows programs of transcription, the generation of RNA components of the translational machinery, and the prevalence of small regulatory RNAs. The 5' ends of primary transcripts of all class of RNA can be differentiated from 5' ends generated by cleavage steps in the processing or degradation of RNA. Studies of other bacterial systems suggest that most 5' ends of ‘secondary’ transcripts generated during processing or degradation will have a 5'-monophosphate group, while the vast majority of nascent transcripts of all classes will be synthesised with a 5'-triphosphate group (Carpousis et al., 2009b, Bechhofer, 2009, Belasco, 2010)

To identify the 5' ends of primary transcripts, one half of the RNA sample was treated with tobacco acid pyrophosphatase (TAP), an enzyme that converts 5'-end triphosphates to monophosphate groups (Breter & Rhoads, 1979), prior to constructing and sequencing cDNA libraries of native 5'-end fragments. An increased number of sequencing reads from a 5' end following TAP treatment is an identifier of a TSS. By combining this differential approach with gRNA-seq, we present at single-nucleotide resolution maps of the primary and secondary transcriptomes of *P. acnes*, and demonstrate their utility in exploring gene regulation. To reduce the number of reads stemming from ribosomal RNAs, we removed much of the 23S and 16S species in our samples using commercially available ‘capture’ oligonucleotides. Reads for these rRNA species were still obtained, but represented only 40% of the total. The
incorporation of a fragmentation step allowed the 5' ends of long as well as short RNAs to be characterised. Sequencing was done using the Illumina Solexa platform. TSSs can be identified without erasing the secondary transcriptome using TEX (Terminator™ 5'-Phosphate-Dependent Exonuclease), which is reported to preferentially degrade transcripts terminating with a 5'-monophosphate group.
4.2 Materials and Methods

4.2.1 Differential RNA-sequencing

The same RNA samples (before and after potassium down-shift in duplicate) used for microarray and gRNA-seq analysis were sent for dRNA-seq, data was generated by vertis Biotechnologie AG (Germany). The service included the construction of cDNA libraries before and after treatment with tobacco acid pyrophosphatase (TAP), and the alignment of RNA sequences to the genome, which was retrieved from NCBI (accession number AE017283). The 5'-sequencing adaptor was ligated to transcripts prior to fragmentation, thereby allowing the 5' ends of both long and short transcript to be detected. RNA samples were enriched for mRNA using MICROBExpress™-Bacteria beads, as described by the manufacturer (Ambion). Illumina Solexa platform was used for the sequencing. Pairs of datasets were compared using M-A scatterplots as described in Chapter 3, except for each A value, we calculated the average ($\mu$) and standard deviation ($\sigma$) of M in a moving window of 5,000 pairs sorted in ascending order of A. Upper and lower envelopes were defined by the equation: $\mu \pm 3\sigma$, and positions outside the envelope recorded, as described previously (Hovatta et al., 2005, Marincs et al., 2006). Details of specific comparisons are provided in the Results section.
4.3 Results

4.3.1 Transcriptional start sites

Transcriptomes of duplicate cultures of *P. acnes* grown as described in Chapter 3 were analysed. The differential approach described here used 8 cDNA libraries; 2 replicates x 2 conditions x 2 treatments (minus or plus TAP treatment). Three to six million reads were obtained for each library and mapped onto the *P. acnes* genome. For each library, for each position in the genome, the number of times it was the first nucleotide of a sequencing read was counted. For each replicate and condition, M-A scatterplots (where M = $\log_2$ (reads plus/minus TAP treatment), and A = $(\log_2$ plus + $\log_2$ minus)/2) revealed a population of values that centred close to an M value of 0, corresponding to sites of processing and degradation, and another with higher M values, corresponding to transcriptional start sites (Figure 4.1). The envelope of the population corresponding to sites of processing and degradation was defined using an established method (see Material and Methods). Nucleotide positions with M values above the envelope that contained sites of processing and degradation were then identified. To increase the power of our analysis, we combined the sequencing results before and after potassium downshift. Positions with M values above the envelope in each of the four experiments (2 duplicates x 2 conditions) were designated positions of transcription initiation, and positions within 8 nt of each other were classified as belonging to the same TSS. With regard to the latter, it is well established that many promoters can initiate transcription at more than one nucleotide position. For the few genes that showed a change in gene expression following potassium downshift, the stringency of the analysis was reduced to the condition under which transcription could be detected most readily. By this approach we identified 4,058 TSSs (Table 51).

The majority of the reads that we obtained by differential RNA-seq (>99.5%) represented processing and degradation sites (PDSs). *P. acnes* encodes three endoribonucleases, RNase E (Ghora & Apirion, 1978), RNase Y (Shahbabian et al., 2009) and RNase III (Robertson et al., 1968), and a dual endonuclease/5' to 3' exonuclease, RNase J (Even et al., 2005, Mathy et al., 1968).
2007), (Table S3) that could account for the large number of PDSs detected in the transcriptome.

Figure 4.1. M-A scatterplots of values from differential RNA-seq. Panels A and B show represent the data for cells cultured without and with a potassium downshift, respectively. Panels C and D, as A and B, except data is for a duplicate pair of cultures. The M values correspond to Log₂ (Plus/Minus) and A values to (Log₂ Plus + Log₂ Minus)/2, where minus and plus refer to the number of reads before and after treatment with TAP. For further details, see Materials and Methods. The red and blue points represent the upper and lower boundary of
the envelope containing sites of processing. The boundaries were defined by sorting the pairs of values in ascending order of A, and then calculating the average ($\mu$) and standard deviation ($\sigma$) of M in a moving window of 5,000 pairs. Upper and lower envelopes were defined by the equation: $\mu \pm 3\sigma$, and positions outside the envelope recorded, as described previously (Hovatta et al., 2005, Marincs et al., 2006).

Next we viewed the positions of the TSSs alongside the results of global RNA-seq, which reveals the 3', as well as 5', boundaries of transcripts and their abundance (Wang et al., 2009b, Marguerat & Bahler, 2010). For each position in the genome, we determined the number of times it was read by global RNA-seq. Viewing the two sets of RNA-seq data side by side revealed that the TSSs we identified belong to all classes of functional RNA. We detected TSSs for mono- and poly-cistronic mRNA, transfer RNA, ribosomal RNA, and ubiquitous small RNAs (Figure 4.2). This included TSSs identified for the pqs locus, which encodes a two-component system thought to be involved in quorum sensing, using standard 5’ RACE (Guan, 2011). Interestingly, the majority of TSSs we identified appeared to be associated with short transcripts of low abundance (< 50 reads, see panel G). The functions of these transcripts are cryptic with a large proportion being found within coding regions on both the sense and antisense strands. Evidence for ‘pervasive’ transcription, a widespread phenomenon in eukaryotes (Jacquier, 2009, Marguerat & Bahler, 2010), is emerging in bacteria (Albrecht et al., 2010, Beaume et al., 2010, Cho et al., 2009, Dornenburg et al., 2010, Filiatrault et al., 2010, Georg et al., 2009, Guell et al., 2009, Jager et al., 2009, Lasa et al., 2011, Liu et al., 2009, Martin et al., 2010b, Mendoza-Vargas et al., 2009, Mitschke et al., 2011, Rasmussen et al., 2009, Sharma et al., 2010, Toledo-Arana et al., 2009, Wurtzel et al., 2010). Of the TSSs we identified, 1106 were associated with step increases in transcription that continued into annotated genes, as illustrated in Figure 5.2, or produced discrete RNAs of high abundance relative to flanking regions (Table S2).
Figure 4.2. TSSs associated with examples of different classes of RNA. Panels A, B, C, D, E, F and G show data corresponding to monocistronic mRNA (PPA0001/0002), polycistronic mRNA (PPA1308-1310), tRNA, rRNA, tmRNA, SRP RNA, and pervasive transcriptional start sites, respectively. The panels are screenshots from the UCSC Microbial Genome Browser (Schneider et al., 2006). In each panel the tracks depict from top to bottom, the position of annotated genes (protein and, as appropriate, RNA coding), the number of times each nucleotide position was the first in sequence reads before and after treatment with TAP (dRNA-seq data), the positions of TSS identified by the analysis of M-A scatterplots (Table S1), and the number of time each position was sequenced following fragmentation of the transcriptome (gRNA-seq). The numbers at the left of RNA-seq tracks indicate the scale of the sequencing reads, while the numbers at the top of each panel indicate the genome position. TSSs in black text were judged by viewing of the gRNA-seq data to be associated with step increases in transcription, while those in blue text were not.
4.3.2 Transcription and maturation of stable RNAs

We find that the 45 tRNAs encoded by the *P. acnes* genome are organised as 41 transcriptional units (data not shown). In stark contrast to what has been found for *B. subtilis* (Dittmar et al., 2004), which along with *E. coli* is one of the main model systems in which tRNA processing has been studied in detail (Hartmann et al., 2009), none of the *P. acnes* tRNA genes are part of the rRNA operons in *P. acnes*, of which there are 3. Another striking difference is that most *P. acnes* tRNA genes are transcribed individually: we only found one example of a tricistronic tRNA operon (Val, GAC; Cys, GCA; Gly, GCC), and two examples of dicistronic operons (Met, CAT; Thr, GGT; and Asp, GTC; Phe, GAA). Thus, co-transcription does not appear to be a major means of regulating stable RNA production in *P. acnes*, unlike the situation in *B. subtilis*. 
Figure 4.3. Location of 3' processing sites for tRNA. Panels A, B, C and D show correspond to tRNAs Gln, TTG (PPA2415), Gly, CCC (PPA2455), Ser, TGA (PPA2461) and Gly, TCC (PPA2438), respectively. For each panel, the tracks depict from top to bottom the position of annotated genes, the average of the dRNA-seq values before and after TAP treatment (combining values for the control and potassium-downshift sample), and values obtained by gRNA-seq. Vertical lines and arrows indicate the location of 3’ side of CCA encoded in the genome and 3’ processing sites, respectively. Remainder of the labelling, as Figure 4.2.
Analysis of our differential RNA-seq data revealed processing sites (i.e. 5’ end sequences of intermediates) within several nucleotides of the annotated 3’ end of 44 of the 45 *P. acnes* tRNAs (for examples, see Figure 4.3). This was a surprise, as just over half of the tRNA genes in *P. acnes* encode a CCA triplet at their 3’ end. In current models of tRNA 3’ processing in eubacteria, the 3’ end of such tRNA is thought to be generated by the action of 3’ exonucleases (Hartmann et al., 2009). We detected processing immediately 3’ to ends to which CCA could be added post-transcriptionally and 3’ to template encoded CCA codons (see panels A and B, respectively). More surprising, the processing sites in approximately half of the tRNA transcribed with the CCA triplet removed this motif (see panel C and D), which is essential for tRNA function. However, the bulk of tRNAs have a CCA motif as detected by gRNA-seq. This apparent paradox can be explained, if following endonucleolytic cleavage another CCA triplet is attached post-transcriptionally. *P. acnes* has a homologue of tRNA nucleotidyltransferase, but to our knowledge this enzymes has only been associated with the attachment of CCA to tRNAs not transcribed with this motif at their 3’ ends (Hartmann et al., 2009). There is not obvious *P. acnes* homologue of tRNase Z (Table S3), the endonuclease that can generate the 3’ end for the post-transcriptional attachment of the CCA in *E. coli*, *B. subtilis* and other bacteria (Hartmann et al., 2009). Thus, another 3’ tRNA processing endonuclease remains to be identified in bacteria. This demonstrates the value of studying fundamental processes in bacteria out with accepted model species. *P. acnes* contains homologues of RNase PH and RNase D, two 3’ to 5’ exonucleases involved in trimming tRNA, and tRNA nucleotidyl transferase (Table S3). Differential RNA-seq also identified mature tRNA 5’ ends (Figure 4.3), which are generated by RNase P (Hartmann et al., 2009), an endonuclease composed of a catalytic RNA and a protein (Table S3).

For each of the three rRNA operons in *P. acnes*, two TSS were identified upstream of the 16S rRNA gene, an arrangement reported previously in *E. coli* (de Boer et al., 1979, Gilbert et al., 1979, Young & Steitz, 1979) and *B. subtilis* (Stewart & Bott, 1983). For each operon, we also identified staggered cleavages in complementary regions that flank mature 16S and 23S rRNA.
and facilitate extensive base-pairing (Figure 4.4). These cleavages are likely mediated by the _P. acnes_ homologue of RNase III (Table S3), which is a well characterised double-stranded-specific endoribonuclease (Nicholson, 2003) with a wide-spread role in the maturation of ribosomal RNA (Deutscher, 2009). In addition to sites of putative RNase III cleavage, we identified sites corresponding to the mature 5′ ends of all three ribosomal RNAs and the mature 3′ end of 16S rRNA. We also identified sites within one or two nucleotides downstream of the mature 3′ end of 23S and 5S rRNA. Thus, the maturation of rRNA in _P. acnes_ appears to make extensive use of endoRNases. All of the sites described above were associated with a step-change in transcript levels. Following endonucleolytic cutting, the mature 3′ ends of _P. acnes_ 23S and 5S rRNA trimming of short 3′ tails. Regarding the generation of the mature 5′ end of 16S rRNA, _P. acnes_ has homologues of both RNase J and RNase E (Table S3), ribonucleases that mediate this function in _B. subtilis_ and _E. coli_, respectively (Deutscher, 2009). Interestingly, we also identified cleavage sites within 16S, 23S and 5S rRNA. These may represent steps in controlling the quality of rRNA (and ribosomes) or preventing rRNA accumulation in excess of ribosomal proteins.
Figure 4.4. Location of ribosomal RNA processing sites. Panel A shows an annotated view of the rRNA cluster. The vertical arrows at the bottom indicate the position of transcriptional start sites and major processing sites referred to in the text, respectively. Predicted RNase III sites are labelled. Short horizontal arrows indicate positions of complementary regions that facilitate extensive base-pairing. Panel B and C show the base-pairing sequences of the 5'UTR and 3'UTR of 16S and 23S rRNA, respectively. The nucleotide position of the RNase III cleavage sites are shown and marked by vertical lines. Remainder of the labelling, as Figure 4.2.
Our approach was also able to identify specific cleavage sites in other classes of RNA. For example, we detected putative RNase III sites within a base-paired region of the 5' leader of the mRNA of pnp mRNA (Figure 4.5), which encodes a 3' to 5' exonucleases (Table S3). Cleavage at the equivalent sites in E. coli has been shown to produces 3' ends that are accessible by PNPase, thereby setting an autoregulatory mechanism that ensures that any increase in PNPase production is only transitory as it leads to increased degradation of pnp mRNA (Jarrige et al., 2001, Robert-Le meur & Portier, 1992, Robert-Le meur & Portier, 1994). This autoregulatory mechanism would appear to be evolutionarily conserved: experimental evidence for its existence in Streptomyces spp. has been obtained (Gatewood et al., 2011).

Figure 4.5. Location of RNA processing sites in 5' UTR of pnp mRNA. Panel A shows the RNA-seq data, while panel B shows the location of the processing sites relative to the secondary structure of the 5' UTR as predicted using Mfold (Zuker, 2003). The labelling of panel A is as Figure 4.2. Panel B, the nucleotide positions are numbered relative to the TSS, while the sites of processing are numbered relative to the genome.
4.3.3 Vegetative promoters

To gain knowledge of vegetative promoters in *P. acnes*, we aligned with the aid of MEME (Bailey *et al.*, 2009) the sequences upstream of 92 TSSs associated with genes of the translational machinery (Figure S1). This revealed a hexanucleotide sequences GnTnG and TAnnnT centred on average –36 and –9 nt, respectively, from the centre of the TSSs (Figure 4.6). These sequences and their relative locations are similar to the consensus reported previously for 'vegetative' promoters of *E. coli* (Harley & Reynolds, 1987, Lisser & Margalit, 1993). Following convention established for *E. coli*, we will refer to the above *P. acnes* sequences as ‘-35’ and ‘-10’ boxes, respectively. The consensus sequences of the equivalent boxes in *E. coli* promoters, TTGACA and TATAAT, are centred on average -33 and -10 nt, respectively, from the centre of the TSSs (Harley & Reynolds, 1987, Lisser & Margalit, 1993). The positioning of the -35 box of *E. coli* closer to the TSS, means that the shared TnG (located in the 5’ half of the *E. coli* box and in the 3’ half of the *P. acnes* box) is on average in the same position relative to the TSS in both organisms. Thus, it appears that the sequence specificity of the housekeeping RNA polymerases in *P. acnes* and *E. coli* retain elements in common despite these organisms diverging over 300 million years ago.
Figure 4.6. The conserved sequences of promoters associated with the translational machinery. Panels A and B are Weblogo representations (Crooks et al., 2004) without and with changing the length of the spacer of individual promoters to maximise the alignment of the -35 box (Figure S1). The combined height of nucleotide symbols shows the level of sequence conservation at a particular position, while the height of individual symbols within a stack of nucleotides indicates the relative frequency at that position. The nucleotide positions are numbered relative to the average position of TSSs. In panel B, this numbering only extends to the point at which gaps were introduced to maximise the alignment.
We next analysed the sequences upstream of the 1106 TSSs associated with step increases in transcription. This revealed that the vast majority had appropriate positioned sequences matching the -10 box consensus. For example, using MEME, we identified 872 that matched the single most common sequence variant (TAanntT). As an aside, this finding reinforces that the dRNA-seq approach described above identifies bona fide transcriptional start sites. Computation predictions of promoters in Propionibacterium and related genera that utilised the promoters identified here as a learning set will be presented elsewhere. Consistent with the analysis of the promoters of rRNA, r-protein and tRNA genes (Figure S1 and 4.6), the overall level of sequence conservation at the -35 position was considerably lower. Nevertheless, promoter sequences were identified that also matched the single most common sequence variant of the -35 box consensus (GnTnTnG). In addition to the promoters of rRNA, r-proteins and tRNA genes, this included elongation factor Tu (PPA1873), hypothetical proteins (PPA0201, 1052, 1403, 1421, 1516, 1680, 1879, 1986), TetR family regulator (PPA0529, 1205), 3-oxoacyl-ACO reductase (PPA1533), alanine dehydrogenase (PPA2274), cytochrom d ubiquinol oxidase subunit I (PPA0176), dihydrolipoamide acyltransferase (PPA0693), fructose-1,6-bisphosphate aldolase (PPA2024), isopentenyl-diphosphate delta isomerise (PPA2115), sodium/hydrogen antiporter (PPA2203), nitric-oxide reductase subunit B (PPA1975), polynucleotide phosphorylase (PPA1471), uridylate kinase (PPA1519) and translation initiation factor IF-2, IF-3 (PPA1493, 1414). Moreover, these promoter were associated with some of the highest transcript levels (data not shown), consistent to the well-established finding that promoters with matches to a consensus tend to be ‘strong’ (Huerta & Collado-Vides, 2003).

4.3.4 Uncovering multiple layers of regulation

The identification of TSSs and promoter sequences alongside our high-resolution transcriptome maps provides a much improved platform for assessing the complexity of gene regulation. This is illustrated here using the P. acnes homologue of NdrR, a transcription factor that controls the expression of ribonucleotide reductases (RNRs) (Borovok et al., 2004, Torrents et al., 2007), and the pqs operon (Guan, 2011). By using MEME to compare sequences -60 to +15 relative to TSSs
mapped for ndrR and genes encoding components of RNRs, we were able to identified probable binding sites for NdrR (referred to here as ndr-boxes). These binding sites overlapped some, but not all of the identified promoters: a pair of ndr-boxes overlapping the distal promoter of two for the ndrRJ operon and a single ndr-box overlapping the ndrAB promoter (Figure 4.7). Moreover, after constructing a position-weight matrix and scanning the entire genome of P. acnes using PREDetector (Hiard et al., 2007) we identified another pair of ndr-box far downstream of the ndrDG promoter. Our analysis shows that the transcription of ndrR and some of its targets are under the control of multiple promoters, only some of which are regulated by NdrR.
See next page for figure legend.
Figure 4.7. Transcription, promoters and cis-regulatory motifs within the ndr operons. Panels A, B, and C correspond to the ndrRJ (PPA1025-1026), ndrAB (PPA2121-2122), and ndrDG (PPA2137-2136) operons, respectively. For each of these panels, the tracks show, from top to bottom, the positions of annotated genes, the position of transcriptional start sites (Table S1) and the gRNA-seq data. The insets in each panel indicate the positions of predicted ndr-boxes and their sequences. Remainder of the labelling are as Figure 4.2.

Perhaps more surprisingly, we also found evidence of post-transcriptional control: much of the transcription of the ndrAB operon appears to terminate before the first structural gene (Figure 4.7). Consistent with this interpretation, the 5' UTR region of ndrAB is annotated as containing a cobalamin riboswitch (Griffiths-Jones et al., 2005), a cis-regulatory element that is widely distributed in the 5' UTRs of cobalamin- (vitamin B12) related genes in eubacteria (Barrick & Breaker, 2007, Franklund & Kadner, 1997, Nahvi et al., 2002, Vitreschak et al., 2003). Interestingly, the 5' UTR of ndrDG, but not the ndrRJ operon is also annotated as containing a cobalamin riboswitch. Furthermore, the activity of the RNR encoded by ndrJ, which is co-transcribed with ndrR, is cobalamin dependent (Nordlund & Reichard, 2006), and we can detect expression of cobalamin biosynthetic genes. Thus, the post-transcriptional control we have identified may bias RNR production towards NdrJ when cobalamin is available.

Our results also lead us to propose that under the anaerobic conditions used for this study NdrR is active. The bulk of the transcription of ndrRJ appears to initiate at the distal promoter, not the proximal promoter overlapped by a pair of ndr-boxes, and we did not detect transcription initiation in the immediate vicinity of the pair of ndr-boxes located upstream of ndrDG. We did detect relatively high levels of transcription from an ndrAB promoter, but this is overlapped by only a single ndr-box. We also speculate that should a promoter exist in the vicinity of the ndr-boxes located upstream of ndrDG inactivation of NdrR will produce a transcript lacking a functional riboswitch, thereby removing the cobalamin regulation. The RNR encoded by the ndrDG operon is thought to function under anaerobic conditions.
The \textit{pqs} locus contains a two-component system that is unusual in two regards. The genes of histidine kinase (HK) and response regulator (RR) are divergently transcribed, and the gene encoding histidine kinase is preceded by a gene predicted to encode an extracellular signalling peptide (EPS) (Figure 4.8). Prior to undertaking the approach described here we had studied the transcription of this locus by 5' RACE and qRT-PCR (Guan, 2011). This revealed single promoters upstream of the EPS and RR genes and suggested that the EPS and HK genes were co-expressed at different stages during batch culture. By comparison, our combined RNA seq approach revealed much more. It not only identified both of the promoters identified by 5' RACE and confirmed that transcription from EPS continues into RR, it identified a second TSS upstream of the EPS gene and identified a small antisense RNA overlapping the 5' end of RR transcript (Figure 4.8). Both of these new elements have now been incorporated into a continuing dissection of the \textit{pqs} locus.
Figure 4.8. **RNA-seq analysis of the pqs operon.** The tracks are as Figure 4.7; remainder of labelling as Figure 4.2. For the insets below the gRNA-seq tracks, large bold fonts indicate transcriptional start sites identified by 5' RACE (Guan, 2011), putative -10 and -35 region are in italic font. The genes of histidine kinase, response regulator and the extracellular signalling peptide correspond to PPA0945, PPA0947, and PPA0946, respectively. The green box marks the boundaries for a potential antisense transcript for the PPA0945-0946 transcriptional unit.
4.3.5 Identification of potential sRNAs

As indicated above, we identified a number of TSSs that were associated with step increases in transcription that did not continue into annotated protein-coding or stable RNA genes, but did produce discrete sRNAs of high abundance relative to the flanking regions. In addition to the ubiquitous bacterial sRNA, 6S RNA, tmRNA, SRP and the RNA component of RNase P (Figure 4.2), this group included 18 examples of what appears to be attenuation, the regulated termination of transcription upstream of structural genes (Henkin & Yanofsky, 2002), 41 examples of sRNAs antisense to the transcripts of protein-coding genes and 28 examples of sRNAs largely encoded within intergenic regions (Table S4). The presence of TSSs upstream of the latter indicates that they are not metastable decay intermediates of an mRNA. Examples of each of the sub-groups are shown in Figure 4.9. Our results indicate that, as is being found increasingly in other bacteria, sRNAs are likely to a major role in the regulation of gene expression in P. acnes.
Figure 4.9. Examples of *P. acnes* small RNAs. Panels A, B, C and D correspond to examples of riboswitch, *cis*-encoded antisense RNA, intergenic sRNA of unknown function, and metastable decay intermediates. For each of these panels, the tracks show, from top to bottom, the positions of annotated genes, the position of transcriptional start sites (Table S1) and the gRNA-seq data. Labelling, as Figure 4.2.
4.3.6 Leaderless mRNAs

While mapping TSSs, we noticed several examples that coincided with start codons for translation (Figure 4.10). This prompted us to gauge the prevalence of leaderless mRNAs in P. acnes. The start codons of protein-coding genes, as annotated in the NCBI database (Pruitt et al., 2007), were collated and their positions mapped against TSSs associated with a step increase in transcription (Table S1). This revealed 50 instances of annotated start codons overlapping TSSs and another 88 where start codons followed TSSs within 10 nt (Table S5). The latter produce a 5' leader that is generally considered too short to recruit ribosomes via the canonical Shine-Dalgarno interaction, which requires base pairing between the leader and a complementary sequence in the 3' end of 16S rRNA (Shine & Dalgarno, 1974, Shine & Dalgarno, 1975). We also identified 15 instances of mRNAs with relatively short 5' leaders (<20 nt) within which we were unable to detect Shine-Dalgarno sequence using RBSfinder (Suzek et al., 2001) (Table S5). From the above, we concluded that in sharp contrast to what has been found for nascent mRNAs in E. coli (Janssen, 1993), translation initiation in the absence of a Shine-Dalgarno interaction appears to be prevalent in P. acnes. Analysis of the ontology of genes associated with ‘leaderless’ mRNA failed to identify enrichment of particular functions. We identified leaderless mRNAs associated with biosynthesis of cofactor (PPA1698, PPA1943), DNA metabolism (PPA2027), energy metabolism (PPA0661, PPA1376), phosphorus compounds (PPA0744), protein synthesis (PPA1344), regulatory functions (PPA1568) and the rest being hypothetical protein or unknown function (PPA0220, PPA0346, PPA0436, PPA1011, PPA1046, PPA1246 and PPA1899). The proportions of AUG, GUG, UUG start codons were 68%, 29% and 3%, respectively. This is similar to the proportion reported for another actinomycete, Streptomyces coelicolor (Vockenhuber et al., 2011). The mechanism by which leaderless mRNA is translated in actinomycetes has not been determined, to our knowledge. Very recently it has been shown that leaderless mRNAs can be generated post-transcriptionally by a stress-induced mRNase that is the toxic component of a toxin-antitoxin system (Vesper et al., 2011). Should such processing also exist in P. acnes, the detection of the corresponding sites would require us to add a phosphorylation step to facilitate the cloning and sequencing of the 5'-hydroxylated fragments that are produced by toxin mRNases.
Figure 4.10. Leaderless mRNAs and transcripts of genes requiring reannotation. Panels A and B correspond to examples of leaderless mRNA, while C and D correspond to genes requiring reannotation. For each of these panels, the tracks show, from top to bottom, the positions of annotated genes, the position of transcriptional start sites (Table S1) and the gRNA-seq data. For C and D, the positions of alternative start codons and associated ribosome binding sites (RBS) are indicated. Labelling, as Figure 4.2.
4.3.7 Re-annotation of protein-coding genes and operon structures

We also noticed that a significant proportion of TSSs associated with significant step increases in transcription were internal to the 5’ half of annotated genes (for examples, see Figure 4.10 panel A) suggesting that the actual gene might be shorter. Consistent with this notion, we have been able to find ribosome binding sites associated with appropriately spaced start codons downstream of many of these TSSs (Table S6) and homologues that lack sequences matching the 5’ end of the original annotation (data not shown). Our combined RNA-seq approach also revealed many examples of operon structures that differ significantly from bioinformatics predictions (for examples, see Figure 4.10, panel B). This was not particular surprising; it is known that even the best bioinformatic approaches are not completely accurate (Chuang et al., 2012). Nevertheless, achieving accurate information on gene and operon structures is essential for gene expression and regulation to be modelled (Salgado et al., 2006) at the level of the whole cell (Karr et al., 2012a, Karr et al.). Our transcriptome approach and data should hasten the achievement of this goal for *P. acnes*. 
4.4 Discussion

Here we describe a differential RNA-seq approach that distinguished sites of transcription initiation without erasing the secondary transcriptome using TEX, an enzyme that in our hands can degrade a substantial proportion of 5'′-triphosphorylated RNA under the conditions recommended by the vendor (Figure S2). With the advent of sequencing techniques that can provide in excess of 100 M reads (e.g. Illumina Solexa) there is now no need to erase the secondary transcriptome in order to detect transcriptional start sites. We simply used TAP (Breter & Rhoads, 1979) to distinguish tri- from mono-phosphorylated 5′ ends. This enzyme was used in earlier differential RNA-seq approaches, but to facilitate the cloning of 5′-fragments remaining after TEX treatment (Vockenhuber et al., 2011, Sharma et al., 2010). Other improvements were to fragment the RNA after the addition of the 5′ adaptor to improve the efficient cloning of 5′ ends from large transcripts, and to combine with a global RNA-seq approach that does not require an amplification step (Mamanova et al., 2010b). The latter allowed us to identify readily the 3′, as well as 5′ boundaries of transcripts. Moreover, by including biological replicates in our differential RNA-seq approach and applying a statistical analysis, we are confident in the assignment of the vast majority of TSSs. In each comparison, only 10,000 of the 5,000,000 pairs of values (plus and minus TAP treatment) were outside the envelope of processing sites. Thus, the probability of a positions being outside the envelope in each of four comparisons is less than 1 in 62500000000. The latter number exceeds the total number of 5′ ends that were identified.

Our approach has already advanced enormously our understanding of gene expression and regulation. We have, for example, identified the positions of thousands of TSS and associated transcriptional units belonging to all classes of functional RNA, mono- and poly-cistronic mRNA, transfer RNA, ribosomal RNA, and ubiquitous small RNAs (Figure 4.2). This alone was identified as an important milestone along the route to understanding the cellular workings of P. acnes (Bruggemann et al., 2004). In addition, from our data we have extracted patterns of stable RNA processing, identified a role for an endonuclease other than tRNase Z in the maturation of tRNA
3 ends (Figure 4.3 and 4.4), mapped sites of mRNA processing (Figure 4.5), identified features of vegetative promoters (Figure 4.6) and potential transcription factor binding sites (Figure 4.7), and discovered functioning riboswitches as well as an abundance of sRNAs (Figure 4.9). We have also shown how knowledge of the above can be used to build models of gene regulation that should inform experimental investigation (Figure 4.7).

One of the surprises of our study was the prevalence of leaderless mRNAs in P. acnes, which is in stark contrast to the situation in E. coli, the main eubacterial system in which the translation of leaderless mRNA has been studied (Moll et al., 2002, Malys & McCarthy, 2011). Indeed, only two examples of E. coli leaderless mRNA have been widely reported, the cl repressor gene of bacteriophage lambda (Walz et al., 1976) and the tetR repressor of transposon Tn1721 (Baumeister et al., 1991). The association between leaderless mRNA and repressors within mobile genetic elements in E. coli has been extended to the repressors of the Rac, e14 and Qin prophages by our own deep RNA-seq analysis of the E. coli transcriptome (unpubl. result). We speculate that some aspect of the translation of these leaderless mRNAs may be important in controlling the mobilisation of the corresponding genetic elements. Recently, it has been shown that stress induces the production of specialised ribosomes that selectively translate a group of mRNAs made leaderless by MazF (Vesper et al., 2011), an endoribonuclease of a toxin-antitoxin (TA) module. Like their mRNA targets, the specialised ribosomes are produced by MazF cleavage, which removes 43 nt from the 3' end of E. coli 16S rRNA (Vesper et al., 2011). Intriguingly, we have mapped a processing site 53 nt from the 3' end of P. acnes 16S rRNA. This raises the possibility that specialised ribosomes, similar to those generated by MazF in E. coli, could mediate much of the translation in P. acnes. However, unlike the situation described for E. coli (Van Etten & Janssen, 1998, O'Donnell & Janssen, 2002, Brock et al., 2008), the translation of leaderless mRNA in P. acnes does not appear to require that the start codon is AUG. As described above, a significant proportion of the leaderless mRNA in P. acnes have GUG (29%) and UUG (3%) in addition to AUG (69%) start codons (Table S5).
Very recently it has been shown that a 5'-terminal monophosphate is required for the efficient translation of leaderless mRNA in *E. coli* (Brock et al., 2008), such ‘decapped’ ends can be produced by RppH (Deana et al., 2008), an RNA pyrophosphohydrolase initially shown to initiate a pathway of mRNA decay in *E. coli* (Celesnik et al., 2007). For the vast majority of 5' ends of leaderless mRNAs in *P. acnes* we could detect a proportion (~10% on average) that was 5' monophosphorylated. Moreover, *P. acnes* has a homologue of RppH (PPA0342). Thus, translation of leaderless mRNAs, and perhaps the initiation of mRNA degradation, in *P. acnes* could be dependent on decapping by an RNA pyrophosphohydrolase. Our study adds to a growing body of evidence that leaderless mRNAs are prevalent outside *E. coli* and its closest relatives and the notion that the mechanism of their translation may represent an ancient milestone in the evolution of gene expression (Moll et al., 2002, Malys & McCarthy, 2011). A gene ontology analysis of leaderless mRNA in *P. acnes* revealed a wide distribution of cellular roles (data not shown). Thus, since the emergence of translation mediated by a Shine-Dalgarno interaction, there does not seem to have been divergence in terms of cellular functions that are dependent on leaderless translation in *P. acnes*. It will be interesting to establish for *P. acnes* whether there is a correlation between leaderless translation and the level of gene expression as measured by protein levels or the response of genes under conditions of stress or both.

Another surprise of our study was the finding that the majority of the TSS we identified were not associated with step increases in transcription that continued into annotated genes or produced discrete RNAs of high abundance relative to flanking regions (Table S1). This may represent ‘pervasive’ transcription, which is widespread in eukaryotes, where it has been shown to have an important role in regulating gene expression (Jacquier, 2009, Marguerat & Bahler, 2010). Evidence for pervasive transcription has already been obtained for several bacteria (Albrecht et al., 2010, Beaume et al., 2010, Cho et al., 2009, Dornenburg et al., 2010, Filiatrault et al., 2010, Georg et al., 2009, Guell et al., 2009, Jager et al., 2009, Lasa et al., 2011, Liu et al., 2009, Martin et al., 2010b, Mendoza-Vargas et al., 2009, Mitschke et al., 2011, Rasmussen et al., 2009, Sharma et al., 2010, Toledo-Arana et al., 2009, Wurtzel et al., 2010). This has largely been in the form of the identification of transcripts antisense to those of
annotated genes. Our study indicates that pervasive transcription in bacteria may stem as much from transcription of the coding strand as the non-coding strand of annotated genes. Of 2930 TSS identified within annotated genes (and not associated with obvious transcription of a flanking gene), 1107 produced transcripts sense to the coding strand. Viewing the positions of all TSS not associated with step increases in transcription that continued into annotated genes or produced discrete RNAs of high abundance, revealed a wide genome distribution, the only bias we have detected so far is for the leading strand of replication. The mapping of TSS to specific sites indicates that the initiation of pervasive transcription is not completely random. Indeed, MEME analysis of sequence upstream of the TSS associated with pervasive transcription identified motif similar to the -10 consensus sequence for *P. acnes* vegetative promoter.

An interesting question is whether or not every region of a bacterial chromosome is transcribed in every cell. In other systems, the abundance of RNase P and tmRNA, ubiquitous sRNAs, have been estimated at ca. 200 (Vioque et al., 1988) and 500 (Chauhan & Apirion, 1989, Lee et al., 1978, Glynn et al., 2007) copies per cell, respectively. For these RNAs, we obtained 45,000 and 120,000 reads, respectively providing an estimate of 225-240 reads per transcript per cell. This estimate appears to be reasonable: it yields 62,500 ribosomes per cell, which is within the range reported for *E. coli*, when applied to the 15 million reads for 5S rRNA. The average number of reads obtained for an mRNA was 125 reads, which corresponds to an average of 1 mRNAs per gene for *P. acnes*. The latter is similar to equivalent number that can be calculated for *E. coli* from its known macromolecular composition. Assuming 70,000 ribosomes per *E. coli* cell, an mRNA content that is 5% of the rRNA content, 4,500 protein coding genes, and an average mRNA length of 1.2 knt (which is 3.8 fold shorter than the combined length of ribosomal RNA) gives on average 3.0 mRNAs per gene (70,000 × 0.05 × 3.8 ÷ 4,500). Immediately downstream (within 50 nt) of TSS thought to be associated with pervasive transcription, we estimate the reads increased on average by 15. This equates to 0.08 transcripts per cell. Thus, it appears that while not every region of the *P. acnes* genome will be represented simultaneously in the transcriptome, every region could be represented for 10s of
minutes during the 6 hour life-cycle of a \textit{P. acnes} cell as it grows and then divides. We are not aware of any experimental evidence to indicate that pervasive transcription in bacteria is no more than a consequence of the broad-sequence specificity of RNA polymerases, which means these enzymes can initiate transcription from sub-optimal sites albeit at reduced frequency. Nevertheless, pervasive transcription maybe of evolutionary significance, allowing the transcription, and thus subsequent selection, of genes acquired horizontally.

The analysis reported here, while producing step changes in our understanding of gene regulation, is far from exhaustive. We hope that our nucleotide-resolution maps will encourage others to search for additional factors controlling gene expression in an organism that is emerging as a significant opportunistic human pathogen, an association that is more than skin deep. For example, our data can be mined to identify \textit{cis}-regulatory signals that control transcriptional termination, initiation from the promoters of genes with shared cellular function, or indeed any other aspect of the life-cycle of RNA such as processing and degradation, and to predict the potential structures and targets of small RNAs. With regard to RNA degradation, we were unable to map a proportion of the RNA sequences (~5%) to the \textit{P. acnes} genome. This may reflect the addition of 3' tails by the \textit{P. acnes} homologue of 3' to 5' exonuclease PNPase, which is known to work in reverse (Mohanty & Kushner, 2000). In systems where they have been studies, bacterial 3' tails facilitate more efficient 3' to 5' degradation (Andrade \textit{et al.}, 2009).
Chapter 5

5 Concluding remarks and future work

The work described in this thesis provides an improved platform from which to study *P. acnes* using functional and comparative genomics. It was successful in establishing culture conditions for *P. acnes* that are sufficiently reproducible (Chapter 2) for specific genetic response(s) to be detected readily at the level of the transcriptome (Chapter 3), and in producing nucleotide-resolution maps of the secondary as well as primary transcriptome of this organism (Chapter 4). The latter was achieved using RNA sequencing protocols, which were either tested for the first time on bacteria (Chapter 3) or refined as part of this work (Chapter 4). From the transcriptome maps, it was possible to detect for the first time for *P. acnes* sites of transcriptional initiation, stable RNA processing and mRNA cleavage as well as the locations of riboswitches, small non-coding RNAs, vegetative promoters, and unannotated genes. In addition, these maps revealed the widespread use of leaderless mRNAs, which may be translated by specialised ribosomes, and the existence of pervasive transcription that is associated with both the sense and antisense strands of coding regions. Combined the above has produced a step change in our knowledge of *P. acnes* gene structure and regulation. Our knowledge is now sufficient that one can start to build meaningful models of gene regulation, as illustrated within this thesis using *ndr* genes (Chapter 4). Better knowledge of gene structure and function will also increase the power of comparative genomics. It is now possible, for example, to include in any comparison sRNA as well as protein-coding genes.

One of our next steps will be to compare the transcriptome of *P. acnes* grown as a biofilm. In addition to adding to our knowledge of gene structure and regulation, it will be interesting to obtain an overview of the cell physiology of biofilms and compare with planktonic growth. Another PhD student in department, Thomas Forth, has developed a graphical tool that allows expression data for metabolic genes to be projected onto known metabolic pathways held within the Kyoto Encyclopaedia of Genes and Genome (KEGG) database (Kanehisa et al., 2012),
thereby providing an overview of metabolism under the growth conditions for which the expression data was collected (Forth, 2012). Analysis of the *P. acnes* transcriptome data for cells grown in batch culture (Chapter 2) using KEGG projector (www.tomforth.co.uk/keggprojector) has been initiated. In addition, it would be interesting to add proteomic data to our analysis of *P. acnes*. There is not always a direct correlation between the transcriptome and proteome. Gene expression is determined not only by the cellular level of transcripts, but by the translatability of the mRNAs and the stability of the proteins. Thus, the physiological state of a cell is more accurately reflected by the proteome. The ability to deduce physiological state of the cell in different growth phases and stress responses from proteomic data was demonstrated in study carried out in *B. subtilis* (Volker & Hecker, 2005). As mentioned in discussion section of Chapter 2, proteome analysis of *phoP* mutant in *S. coelicolor* revealed the remodelling of its metabolism to utilise gluconeogenesis to balance the availability of phosphate and correct the imbalance of redox potential (Thomas et al., 2012).

While much can be learned about organisms by studying them in the laboratory, the ultimate goal is to study them in their natural environment, or under conditions as close to their natural environment as possible. For example, it would be interesting to know what genes are expressed (and are thus likely to function) when *P. acnes* is present within hair follicles both health and associated with acne vulgaris. Is the expression of some genes associated with inflammation? The difficulty of such analyses is that it requires transcriptome analyses to be carried out using small amount of biomass. The numbers of *P. acnes* found in the hair follicles are can reach ~$10^7$ (Bojar 2004). This still corresponds to an amount of RNA that is 100 fold lower than that analysed here by RNA-seq. However, recent advances in sequencing technology have made possible single-cell transcriptome sequencing (CEL-Seq) (Tang et al., 2009). The technique starts with reverse transcription of the RNA using a primer that contains at its 5’ end a T7 polymerase promoter. This promoter on the cDNA can then be used to generate transcripts that are sense to the original transcript. The transcripts generated by this linear amplification step are then fragmented and sequenced as per standard RNA-seq.
methodologies. CEL-seq has been shown to be reproducible using different eukaryotic cell types (Hashimshony et al., 2012). Transcriptome sequencing of a single bacterium may be possible using this technique by adding a polyA tail on the 3’ end bacterial mRNA. To my knowledge, single cell transcriptome sequencing on prokaryotic platform has not been reported.

Amplification of the single cell bacterial transcriptome and subsequent analysis using microarray has been published (Kang et al., 2011). Briefly, short cDNA is synthesised from RNA samples using random hexamers. The cDNA is then circularised using DNA ligase and then amplified isothermally using φ29 DNA polymerase and random hexamer (Hutchison et al., 2005). Up to 35 μg of cDNA was generated from a single cell, it was shown that the amplification of cDNA using this technique was able to detect 95% of the transcripts compared to non-amplified sample from a pool of bacteria using microarray (Kang et al., 2011). Thus, technically it is now possible to analyse the transcription profiles of P. acnes colonising hair follicles. The transcriptome amplification technique currently limits to detection of expression by microarray as it does not yet offer strand specificity. My speculation would be that soon sequencing of single bacterium transcription will be made available.

It should also be remembered that P. acnes is only part of a microbial community that colonise the hair follicle. Other members include Staphylococcus epidermidis, and other propionibacteria, it is accepted that P. acnes is not the sole agent that contribute to acne vulgaris (Holland et al., 1978, Cove et al., 2006). Thus, any transcriptome profile of P. acnes in hair follicles should be accompanied by a metagenomic analysis (Grice et al., 2008). This would provide reference genomes on which to map transcriptome of the microbial community, meta-transcriptome. By studying the entire microbiota of hair follicles, association with acnes vulgaris are more likely to be found. Studies of the microbiota of the human gut have already revealed that flora composition can be affected by diets, and in turn affect the susceptibility of the host to diseases (Claesson et al., 2012). In type-2 diabetes patients, genes associated with methane metabolism, membrane associated sugar transport, branched-chain amino acid transport, drug resistance, and oxidative stress were expressed in abundance by members of clostridia, bacteroides and E.
coli in their gut flora. This is in contrast to what was observed in the control (health) group where genes associated with butyrate biosynthesis, metabolism of cofactors and vitamins were expressed by members of flora including clostridiales and faecalibacteria (Qin et al., 2012). The loss of butyrate production in type-2 diabetes patients correlated with increase in opportunistic pathogen found in the gut, this reinforced the hypothesis that the role of flora contribute to the susceptibility of disease (Qin et al., 2012). Host (gender, age and disease) and environmental factor (clothing, hygiene, and lifestyle) no doubt affect the skin microbiota (Grice & Segre, 2011). Similar type of analysis of the metagenomic profile of skin microbiota to acne vulgaris will give broader view of the population involved in the disease process.

While transcriptomics will undoubtedly identify P. acnes genes that function (as judge by their expression) during growth a biofilm and as part of the microbiota of the skin, there is a yet no straightforward method for knocking out genes in P. acnes to confirm and study their importance to a particular process. While two publications report the disruption of genes in P. acnes via the delivery of cassettes on suicide plasmids by electroporation (Cheong et al., 2008, Sorensen et al., 2010), it has not been possible despite considerable effort to establish this technique in Leeds (John Wright and Keith Holland, pers. comm.). That others have not cited these papers suggests that our experience might not be unique. As homologous recombination is ubiquitous in bacteria (Vos, 2009), it seems likely that the delivery of gene-disruption cassettes is the limiting step. Another possible means of delivering disruption cassettes is conjugation. Cosmids containing genes disrupted in E. coli have been transferred efficiently into various Streptomyces species (Gust et al., 2004). The cosmids were engineered to contain the origin of transfer, plasmid RP4 (Matsushima et al., 1994), which is highly promiscuous (Furste et al., 1989). Plasmids with this origin of transfer have even been transferred between E. coli and the yeast Saccharomyces cerevisiae. Ideally, disruption cassettes should be delivered via cosmids to maximise the length of flanking sequences and thus the possibility of the cassette replacing the target gene in the chromosome. However, as far as we are aware, a cosmid library of P. acnes DNA has not been constructed. Genes of interest to the laboratory include a putative quorum-sensing (PPA0945-0947), which includes a secreted signal peptide
that upon reaching a critical threshold is thought to be activate an associated histidine kinase that then activates an associated response regulator (Guan, 2011). Ultimately the *P. acnes* research community can aim towards generating a mutant library similar to the Keio collection for *E. coli* (Baba *et al.*, 2006).

To determine the consequences of gene disruptions an infection model is required. A skin equivalent model, as described in Chapter 1, consists of a dermal matrix of fibrin containing fibroblasts is seeded with human keratinocyte to generate a stratified epidermis has been used to characterise host innate immune response upon inoculation of skin pathogens (Holland *et al.*, 2009, Holland *et al.*, 2008). Animal model has used to determine the effect of treatment upon *P. acnes* infection. Mouse ear were inoculation with $3 \times 10^6$ of *P. acnes* or sterile saline as control. The infected mouse ear was examined before and after the treatment to determine the treatment efficiency (Fan *et al.*, 2012).

These models could also be used as the basis of functional genomic approaches that are based on creating a library of random mutants and then determine those that are still able to survive and colonise. This approach has been used to identify genes of *Salmonella typhimurium* that were required for this organism, which causes typhoid, to survive in a mouse (Chaudhuri *et al.*, 2009). The mutants were created using barcoded transposons. Barcodes present in the starting pool of mutants, but not recovered from the spleen of the infected mouse, marked bacterial genes required for survival and colonisation (Mazurkiewicz *et al.*, 2006). The development of a similar system, which is called signature-tagged mutagenesis, for *P. acnes* could produce a step change in our understanding of the bacterial factors required for colonisation of the skin and perhaps even *acnes vulgaris*. The latter would require the comparison of barcodes isolated from health follicles and those associated with disease. The prerequisite for this technique is a transposon that can integrate into *P. acnes* genome or an existing library of mutant.
6 Supplementary figures and tables

Table S1. Transcriptional start sites identified for *P. acnes*. The corresponding values where above the upper envelope boundary in all of 4 experiments (see Figure 4.1). 'Nucleotide positions within 8 nt of each other were classified as belonging to the same TSS. *Whether a TSS was associated with a step increase in transcription of was judged by manual inspection of the global RNA-seq data. See excel file.
Table S2. Transcription start sites associated with discrete RNAs that do not cover annotated genes. This is a subset of the data shown in Table S1. *Transcripts judged to correspond to the 5' UTR of mRNAs. *Possible cis-encoded antisense RNA regulators of mRNA.

<table>
<thead>
<tr>
<th>Forward strand</th>
<th>Reserve strand</th>
</tr>
</thead>
<tbody>
<tr>
<td>16650</td>
<td>1129092</td>
</tr>
<tr>
<td>69338-40</td>
<td>1220499</td>
</tr>
<tr>
<td>81851</td>
<td>1224843</td>
</tr>
<tr>
<td>95813</td>
<td>1248645</td>
</tr>
<tr>
<td>133015</td>
<td>1565719</td>
</tr>
<tr>
<td>145183</td>
<td>1566109</td>
</tr>
<tr>
<td>186527</td>
<td>1614866</td>
</tr>
<tr>
<td>199776</td>
<td>1654717</td>
</tr>
<tr>
<td>218860</td>
<td>1669453</td>
</tr>
<tr>
<td>252385</td>
<td>1674593</td>
</tr>
<tr>
<td>262666-69</td>
<td>1725349</td>
</tr>
<tr>
<td>267845-49</td>
<td>1920360</td>
</tr>
<tr>
<td>326577</td>
<td>1933084</td>
</tr>
<tr>
<td>371740</td>
<td>1948896</td>
</tr>
<tr>
<td>382217</td>
<td>1990644</td>
</tr>
<tr>
<td>396972</td>
<td>2066299</td>
</tr>
<tr>
<td>420176</td>
<td>2153965</td>
</tr>
<tr>
<td>476077</td>
<td>2174792</td>
</tr>
<tr>
<td>551864</td>
<td>2269893</td>
</tr>
<tr>
<td>656880-83</td>
<td>2280058</td>
</tr>
<tr>
<td>720186</td>
<td>2294577</td>
</tr>
<tr>
<td>789397</td>
<td>2302239</td>
</tr>
<tr>
<td>892305</td>
<td>2345611</td>
</tr>
<tr>
<td>934712</td>
<td>2386888</td>
</tr>
<tr>
<td>947536</td>
<td>2441486</td>
</tr>
<tr>
<td>951357</td>
<td>2505019</td>
</tr>
<tr>
<td>981381</td>
<td>2538808</td>
</tr>
<tr>
<td>996061</td>
<td>2551263</td>
</tr>
<tr>
<td>1078903</td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>Description</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>RNase E/G (NTD)</td>
<td>Single-strand-specific endoRNase involved in RNA degradation and the processing of stable RNAs. PPA0826 has extension on N-terminus</td>
</tr>
<tr>
<td>RNase III</td>
<td>Double-strand-specific endoRNase involved in processing of rRNA and mRNA. May initiate the cleavage of some mRNAs</td>
</tr>
<tr>
<td>RNase P (RNA component)</td>
<td>EndoRNase that processes 5' end of tRNA. Also processes its own catalytic RNA and cuts some mRNAs</td>
</tr>
<tr>
<td>RppH</td>
<td>RNA pyrophosphohydrolase that initiates degradation of some mRNA by hydrolysis of the 5'-triphosphate end</td>
</tr>
<tr>
<td>RNase J</td>
<td>RNase with dual endo and 5'→3' exo activity, has roles in the degradation of specific structural mRNAs. Does not appear to be critical for mRNA degradation in B. subtilis, not found in E. coli</td>
</tr>
<tr>
<td>RNase Y</td>
<td>Endonuclease involved in the degradation of mRNA in B. subtilis, not found in E. coli</td>
</tr>
<tr>
<td>PNPase</td>
<td>3' to 5' exoRNase and 3'-terminal and oligonucleotide polymerase. Functions in the degradation of various mRNAs and tRNA maturation</td>
</tr>
<tr>
<td>oligoRNase</td>
<td>Processive 3'-to-5' exoRNase specific for short oligoribonucleotides. Final enzyme in degraded RNAs to mononucleotides</td>
</tr>
<tr>
<td>poly(A) polymerase or tRNA nucleotidyl-transferase</td>
<td>Responsible for oligoadenylolation of 3' ends of RNA molecules</td>
</tr>
<tr>
<td>RNase PH</td>
<td>3' to 5' exoRNase involved in 3' trimming of tRNAs.</td>
</tr>
<tr>
<td>RNase D</td>
<td>3'-5' exoRNase involved in the 3' processing of various stable RNA molecules.</td>
</tr>
<tr>
<td>tRNase Z</td>
<td>EndoRNase that can generate the mature 3' end of tRNA</td>
</tr>
<tr>
<td>nanoRNase</td>
<td>Functionally equiva lent to oligoRNase</td>
</tr>
<tr>
<td>RNase BN</td>
<td>3' to 5' exoRNase involved in 3' trimming of tRNAs as well as various short unstructured RNAs</td>
</tr>
<tr>
<td>RNase T</td>
<td>3' to 5' exoRNase responsible for 3' trimming of many stable RNAs, including tRNAs and 5S rRNA. Can compensate for lack of other 3' to 5' exonucleases in tRNA maturation</td>
</tr>
<tr>
<td>RNase II and RNase R</td>
<td>3' to 5' exoRNase that cleaves RNA from the 3' end to produce ribonucleoside 5'-monophosphates</td>
</tr>
<tr>
<td>nanoRNase</td>
<td>Functionally equivalent to oligoRNase</td>
</tr>
</tbody>
</table>
Table S4. List of annotated and novel sRNAs in *P. acnes*.

<table>
<thead>
<tr>
<th>Antisense</th>
<th>Riboswitch</th>
<th>Intergenic region</th>
</tr>
</thead>
<tbody>
<tr>
<td>25506 +</td>
<td>981518 -</td>
<td>16650 +</td>
</tr>
<tr>
<td>48200 -</td>
<td>1062051 -</td>
<td>33231 +</td>
</tr>
<tr>
<td>54282 +</td>
<td>1202004 -</td>
<td>95813 +</td>
</tr>
<tr>
<td>107414 +</td>
<td>1251316 -</td>
<td>199776 +</td>
</tr>
<tr>
<td>226602 +</td>
<td>1543535 +</td>
<td>218860 +</td>
</tr>
<tr>
<td>234105 +</td>
<td>1573376 +</td>
<td>326577 +</td>
</tr>
<tr>
<td>262666 +</td>
<td>1625671 -</td>
<td>476077 +</td>
</tr>
<tr>
<td>350628 +</td>
<td>1731448 +</td>
<td>578506 -</td>
</tr>
<tr>
<td>419298 +</td>
<td>1747014 -</td>
<td>1129093 +</td>
</tr>
<tr>
<td>429137 -</td>
<td>1892720 +</td>
<td>1163170 -</td>
</tr>
<tr>
<td>447936 -</td>
<td>1902900 +</td>
<td>1220499 +</td>
</tr>
<tr>
<td>515180 -</td>
<td>2067110 -</td>
<td>1719382 -</td>
</tr>
<tr>
<td>523981 +</td>
<td>2154770 -</td>
<td>1913792 -</td>
</tr>
<tr>
<td>525730 +</td>
<td>2154776 -</td>
<td>1920361 +</td>
</tr>
<tr>
<td>586175 -</td>
<td>2286406 -</td>
<td>1933084 +</td>
</tr>
<tr>
<td>601083 -</td>
<td>2286407 -</td>
<td>1972785 -</td>
</tr>
<tr>
<td>601251 -</td>
<td>2338912 +</td>
<td>2294577 +</td>
</tr>
<tr>
<td>662829 -</td>
<td>2354014 +</td>
<td>2538808 +</td>
</tr>
<tr>
<td>690031 +</td>
<td>2354015 +</td>
<td></td>
</tr>
<tr>
<td>776546 +</td>
<td>2386888 +</td>
<td></td>
</tr>
<tr>
<td>885274 +</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(+/-) indicates sRNA located on the forward and reverse strand, respectively. *List is not exhaustive, functions are putative and await experimental investigation. Although represented by single nucleotide positions, 3’ boundaries can be broad.*
Table S5. List of leaderless mRNAs.

<table>
<thead>
<tr>
<th>Leaderless mRNA TSS</th>
<th>Strand</th>
<th>Distance from start codon</th>
<th>Gene</th>
<th>Leaderless mRNA TSS</th>
<th>Strand</th>
<th>Distance from start codon</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>85492</td>
<td>+</td>
<td>1</td>
<td>PPA0079</td>
<td>1332639</td>
<td>+</td>
<td>2</td>
<td>PPA1225</td>
</tr>
<tr>
<td>219005</td>
<td>+</td>
<td>0</td>
<td>PPA0180</td>
<td>1335132</td>
<td>+</td>
<td>0</td>
<td>PPA1227</td>
</tr>
<tr>
<td>264693</td>
<td>+</td>
<td>8</td>
<td>PPA0213</td>
<td>1439333</td>
<td>+</td>
<td>0</td>
<td>PPA1323</td>
</tr>
<tr>
<td>275563*</td>
<td>+</td>
<td>0</td>
<td>PPA0220</td>
<td>1486763</td>
<td>+</td>
<td>0</td>
<td>PPA1364</td>
</tr>
<tr>
<td>301857</td>
<td>+</td>
<td>18</td>
<td>PPA0248</td>
<td>1580884</td>
<td>+</td>
<td>0</td>
<td>PPA1465</td>
</tr>
<tr>
<td>310295</td>
<td>+</td>
<td>20</td>
<td>PPA0257</td>
<td>1614866</td>
<td>+</td>
<td>0</td>
<td>PPA1496</td>
</tr>
<tr>
<td>312708</td>
<td>+</td>
<td>0</td>
<td>PPA0259</td>
<td>1677186</td>
<td>+</td>
<td>0</td>
<td>PPA1557</td>
</tr>
<tr>
<td>332153</td>
<td>+</td>
<td>0</td>
<td>PPA0282</td>
<td>1687656*</td>
<td>+</td>
<td>0</td>
<td>PPA1568</td>
</tr>
<tr>
<td>357872</td>
<td>+</td>
<td>15</td>
<td>PPA0307</td>
<td>1789002</td>
<td>+</td>
<td>10</td>
<td>PPA1642</td>
</tr>
<tr>
<td>383235</td>
<td>+</td>
<td>0</td>
<td>PPA0333</td>
<td>1851527</td>
<td>+</td>
<td>0</td>
<td>PPA1699</td>
</tr>
<tr>
<td>396972</td>
<td>+</td>
<td>9</td>
<td>PPA0345</td>
<td>1875582</td>
<td>+</td>
<td>0</td>
<td>PPA1722</td>
</tr>
<tr>
<td>453574</td>
<td>+</td>
<td>0</td>
<td>PPA0408</td>
<td>1924939</td>
<td>+</td>
<td>0</td>
<td>PPA1762</td>
</tr>
<tr>
<td>544192</td>
<td>+</td>
<td>0</td>
<td>PPA0494</td>
<td>2070668</td>
<td>+</td>
<td>0</td>
<td>PPA1909</td>
</tr>
<tr>
<td>565188</td>
<td>+</td>
<td>0</td>
<td>PPA0513</td>
<td>2103935*</td>
<td>+</td>
<td>1</td>
<td>PPA1943</td>
</tr>
<tr>
<td>581080</td>
<td>+</td>
<td>0</td>
<td>PPA0527</td>
<td>2113770</td>
<td>+</td>
<td>0</td>
<td>PPA1953</td>
</tr>
<tr>
<td>619022</td>
<td>+</td>
<td>6</td>
<td>PPA0559</td>
<td>2124998</td>
<td>+</td>
<td>20</td>
<td>PPA1962</td>
</tr>
<tr>
<td>620292</td>
<td>+</td>
<td>0</td>
<td>PPA0560</td>
<td>2200062*</td>
<td>+</td>
<td>2</td>
<td>PPA2027</td>
</tr>
<tr>
<td>632907</td>
<td>+</td>
<td>0</td>
<td>PPA0572</td>
<td>2262758</td>
<td>+</td>
<td>13</td>
<td>PPA2088</td>
</tr>
<tr>
<td>636616</td>
<td>+</td>
<td>0</td>
<td>PPA0575</td>
<td>2270276</td>
<td>+</td>
<td>12</td>
<td>PPA2095</td>
</tr>
<tr>
<td>649446</td>
<td>+</td>
<td>0</td>
<td>PPA0588</td>
<td>2285027</td>
<td>+</td>
<td>0</td>
<td>PPA2111</td>
</tr>
<tr>
<td>651011</td>
<td>+</td>
<td>0</td>
<td>PPA0590</td>
<td>2343229</td>
<td>+</td>
<td>0</td>
<td>PPA2163</td>
</tr>
<tr>
<td>701664</td>
<td>+</td>
<td>0</td>
<td>PPA0636</td>
<td>2359454</td>
<td>+</td>
<td>0</td>
<td>PPA2177</td>
</tr>
<tr>
<td>707657</td>
<td>+</td>
<td>0</td>
<td>PPA0643</td>
<td>2365576</td>
<td>+</td>
<td>0</td>
<td>PPA2184</td>
</tr>
<tr>
<td>712934</td>
<td>+</td>
<td>20</td>
<td>PPA0648</td>
<td>2367840</td>
<td>+</td>
<td>1</td>
<td>PPA2187</td>
</tr>
<tr>
<td>719343</td>
<td>+</td>
<td>0</td>
<td>PPA0653</td>
<td>2380556</td>
<td>+</td>
<td>0</td>
<td>PPA2200</td>
</tr>
<tr>
<td>720186</td>
<td>+</td>
<td>12</td>
<td>PPA0656</td>
<td>2383635</td>
<td>+</td>
<td>0</td>
<td>PPA2202</td>
</tr>
<tr>
<td>744036</td>
<td>+</td>
<td>0</td>
<td>PPA0676</td>
<td>2388012</td>
<td>+</td>
<td>0</td>
<td>PPA2205</td>
</tr>
<tr>
<td>795728</td>
<td>+</td>
<td>13</td>
<td>PPA0724</td>
<td>2420039</td>
<td>+</td>
<td>0</td>
<td>PPA2236</td>
</tr>
<tr>
<td>802023</td>
<td>+</td>
<td>1</td>
<td>PPA0730</td>
<td>2430005</td>
<td>+</td>
<td>0</td>
<td>PPA2246</td>
</tr>
<tr>
<td>843812</td>
<td>+</td>
<td>0</td>
<td>PPA0769</td>
<td>2441486</td>
<td>+</td>
<td>0</td>
<td>PPA2257</td>
</tr>
<tr>
<td>977946</td>
<td>+</td>
<td>2</td>
<td>PPA0897</td>
<td>2453150</td>
<td>+</td>
<td>18</td>
<td>PPA2267</td>
</tr>
<tr>
<td>1007244</td>
<td>+</td>
<td>0</td>
<td>PPA0924</td>
<td>2484814</td>
<td>+</td>
<td>0</td>
<td>PPA2289</td>
</tr>
<tr>
<td>1025484</td>
<td>+</td>
<td>0</td>
<td>PPA0944</td>
<td>2486987</td>
<td>+</td>
<td>0</td>
<td>PPA2292</td>
</tr>
<tr>
<td>1042580</td>
<td>+</td>
<td>0</td>
<td>PPA0959</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1094775*</td>
<td>+</td>
<td>15</td>
<td>PPA1011</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1101306</td>
<td>+</td>
<td>3</td>
<td>PPA1017</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1131717</td>
<td>+</td>
<td>0</td>
<td>PPA1040</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1152133</td>
<td>+</td>
<td>2</td>
<td>PPA1060</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leaderless mRNA TSS</td>
<td>Distance from start</td>
<td>Gene</td>
<td>Leaderless mRNA TSS</td>
<td>Distance from start</td>
<td>Gene</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>---------------------</td>
<td>--------</td>
<td>---------------------</td>
<td>---------------------</td>
<td>--------</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Strand</td>
<td>codon</td>
<td></td>
<td>Strand</td>
<td>codon</td>
<td></td>
<td></td>
</tr>
<tr>
<td>157289</td>
<td>-</td>
<td>1</td>
<td>PPA0130</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>160753</td>
<td>-</td>
<td>1</td>
<td>PPA0133</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>213202</td>
<td>-</td>
<td>1</td>
<td>PPA0176</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>307514</td>
<td>-</td>
<td>19</td>
<td>PPA0252</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>310324</td>
<td>-</td>
<td>1</td>
<td>PPA0256</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>366974</td>
<td>-</td>
<td>1</td>
<td>PPA0314</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>380907</td>
<td>-</td>
<td>1</td>
<td>PPA0330</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>398318*</td>
<td>-</td>
<td>1</td>
<td>PPA0346</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>406992</td>
<td>-</td>
<td>17</td>
<td>PPA0357</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>412518</td>
<td>-</td>
<td>1</td>
<td>PPA0364</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>480388*</td>
<td>-</td>
<td>1</td>
<td>PPA0436</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>527105</td>
<td>-</td>
<td>1</td>
<td>PPA0475</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>578338</td>
<td>-</td>
<td>1</td>
<td>PPA0523</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>583253</td>
<td>-</td>
<td>1</td>
<td>PPA0529</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>621885</td>
<td>-</td>
<td>1</td>
<td>PPA0561</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>625488</td>
<td>-</td>
<td>1</td>
<td>PPA0564</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>629901</td>
<td>-</td>
<td>1</td>
<td>PPA0569</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>635103</td>
<td>-</td>
<td>1</td>
<td>PPA0573</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>664151</td>
<td>-</td>
<td>18</td>
<td>PPA0599</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>689518</td>
<td>-</td>
<td>1</td>
<td>PPA0623</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>696913</td>
<td>-</td>
<td>1</td>
<td>PPA0630</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>706656</td>
<td>-</td>
<td>1</td>
<td>PPA0641</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>712849</td>
<td>-</td>
<td>1</td>
<td>PPA0647</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>712866</td>
<td>-</td>
<td>18</td>
<td>PPA0647</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>727624*</td>
<td>-</td>
<td>1</td>
<td>PPA0661</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>728897</td>
<td>-</td>
<td>1</td>
<td>PPA0663</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>735523</td>
<td>-</td>
<td>3</td>
<td>PPA0667</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>759849</td>
<td>-</td>
<td>1</td>
<td>PPA0689</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>768211</td>
<td>-</td>
<td>19</td>
<td>PPA0694</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>781355</td>
<td>-</td>
<td>1</td>
<td>PPA0708</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>817762*</td>
<td>-</td>
<td>1</td>
<td>PPA0744</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>924594</td>
<td>-</td>
<td>1</td>
<td>PPA0845</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1069245</td>
<td>-</td>
<td>1</td>
<td>PPA0986</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1141827*</td>
<td>-</td>
<td>1</td>
<td>PPA1046</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1146258</td>
<td>-</td>
<td>1</td>
<td>PPA1052</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1160415</td>
<td>-</td>
<td>13</td>
<td>PPA1064</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1196203</td>
<td>-</td>
<td>1</td>
<td>PPA1104</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1198468</td>
<td>-</td>
<td>1</td>
<td>PPA1105</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table S5 continued
Table S5 continued

<table>
<thead>
<tr>
<th>Leaderless mRNA TSS</th>
<th>Strand</th>
<th>Distance from start codon</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>2174714</td>
<td>-</td>
<td>1</td>
<td>PPA2002</td>
</tr>
<tr>
<td>2181571</td>
<td>-</td>
<td>1</td>
<td>PPA2008</td>
</tr>
<tr>
<td>2187838</td>
<td>-</td>
<td>1</td>
<td>PPA2015</td>
</tr>
<tr>
<td>2289917</td>
<td>-</td>
<td>1</td>
<td>PPA2115</td>
</tr>
<tr>
<td>2305442</td>
<td>-</td>
<td>1</td>
<td>PPA2128</td>
</tr>
<tr>
<td>2311086</td>
<td>-</td>
<td>1</td>
<td>PPA2133</td>
</tr>
<tr>
<td>2410592</td>
<td>-</td>
<td>1</td>
<td>PPA2225</td>
</tr>
<tr>
<td>2419928</td>
<td>-</td>
<td>1</td>
<td>PPA2235</td>
</tr>
<tr>
<td>2427001</td>
<td>-</td>
<td>13</td>
<td>PPA2243</td>
</tr>
<tr>
<td>2436830</td>
<td>-</td>
<td>0</td>
<td>PPA2251</td>
</tr>
<tr>
<td>2493898</td>
<td>-</td>
<td>1</td>
<td>PPA2297</td>
</tr>
</tbody>
</table>

*TSSs where no Shine-Dalgarno sequence was found using RBSfinder.
Table S6. List of genes requiring reannotation.

<table>
<thead>
<tr>
<th>Genes requiring reannotation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PPA0048 PPA0049 PPA0051 PPA0120 PPA0139 PPA0172 PPA0199 PPA0220 PPA0228 PPA0259 PPA0275 PPA0344 PPA0353 PPA0358 PPA0360 PPA0384 PPA0408 PPA0417 PPA0482 PPA0494 PPA0496 PPA0527 PPA0671 PPA0676 PPA0690 PPA0769 PPA0774 PPA0781</td>
<td>PPA0885 PPA0892 PPA0924 PPA0934 PPA0935 PPA0941 PPA0946 PPA0949 PPA0965 PPA1024 PPA1102 PPA1106 PPA1121 PPA1167 PPA1171 PPA1226 PPA1267 PPA1267 PPA1310 PPA1314 PPA1329 PPA1350 PPA1364 PPA1410 PPA1416 PPA1474 PPA1712 PPA1734 PPA1764 PPA1777 PPA1794 PPA1808 PPA1871 PPA1906 PPA1913 PPA1947 PPA1953 PPA1954 PPA1983 PPA1984 PPA1995 PPA1996 PPA2126 PPA2151 PPA2166 PPA2183 PPA2202 PPA2251 PPA2270 PPA2292 PPA2299 PPA2314 PPA2315 PPA2341</td>
</tr>
</tbody>
</table>
Table S7. Composition of Holland defined medium

Solution I $^{a,b}$

<table>
<thead>
<tr>
<th>Components</th>
<th>Per 100ml (mg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-Alanine</td>
<td>10</td>
</tr>
<tr>
<td>L-Arginine</td>
<td>20</td>
</tr>
<tr>
<td>L-Asparagine</td>
<td>10</td>
</tr>
<tr>
<td>L-Aspartic acid</td>
<td>20</td>
</tr>
<tr>
<td>L-Glutamic acid</td>
<td>50</td>
</tr>
<tr>
<td>Glycine</td>
<td>10</td>
</tr>
<tr>
<td>L-Histidine</td>
<td>10</td>
</tr>
<tr>
<td>L-Isoleucine</td>
<td>10</td>
</tr>
<tr>
<td>L-Hydroxyproline</td>
<td>10</td>
</tr>
<tr>
<td>L-Cysteine</td>
<td>10</td>
</tr>
<tr>
<td>L-Lysine</td>
<td>10</td>
</tr>
<tr>
<td>L-Leucine</td>
<td>10</td>
</tr>
<tr>
<td>L-Phenylalanine</td>
<td>10</td>
</tr>
<tr>
<td>L-Methionine</td>
<td>10</td>
</tr>
<tr>
<td>L-Proline</td>
<td>10</td>
</tr>
<tr>
<td>L-Serine</td>
<td>20</td>
</tr>
<tr>
<td>L-Threonine</td>
<td>20</td>
</tr>
<tr>
<td>L-Tryptophan</td>
<td>10</td>
</tr>
<tr>
<td>L-Tyrosin</td>
<td>10</td>
</tr>
<tr>
<td>L-Valine</td>
<td>10</td>
</tr>
<tr>
<td>Glucose</td>
<td>500</td>
</tr>
<tr>
<td>PIPES</td>
<td>1000</td>
</tr>
<tr>
<td>K$_2$HPO$_4$</td>
<td>350</td>
</tr>
<tr>
<td>KH$_2$PO$_4$</td>
<td>350</td>
</tr>
<tr>
<td>MgSO$_4$·7H$_2$O</td>
<td>20</td>
</tr>
<tr>
<td>NaCl</td>
<td>200</td>
</tr>
<tr>
<td>Sodium citrate</td>
<td>50</td>
</tr>
<tr>
<td>(NH$_4$)$_2$SO$_4$</td>
<td>150</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mineral salt components$^c$</th>
<th>Concentration (g/100ml)</th>
<th>Vitamin components$^d$</th>
<th>Concentration (g/100ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FeSO$_4$·7H$_2$O</td>
<td>1.0</td>
<td>Pyridoxine HCl</td>
<td>1.25</td>
</tr>
<tr>
<td>MnSO$_4$·4H$_2$O</td>
<td>1.0</td>
<td>Nicotinic acid</td>
<td>0.25</td>
</tr>
<tr>
<td>CaCl$_2$·2H$_2$O</td>
<td>1.0</td>
<td>Pantothenic acid</td>
<td>0.125</td>
</tr>
<tr>
<td>ZnCl$_2$</td>
<td>0.25</td>
<td>Biotin</td>
<td>0.015</td>
</tr>
<tr>
<td>CoCl$_2$·6H$_2$O</td>
<td>0.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CuSO$_4$·5H$_2$O</td>
<td>0.05</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^+$ - Mineral salts were dissolved in 1.0M HCl.
$^\dagger$ - Vitamins were dissolved in dH$_2$O

$^a$ - Tween-80 added at 0.1% [v/v] of Solution I

$^b$ - final pH of HSM is 5.6

$^c$ - Mineral salt and vitamin solutions added at 0.2% [v/v] of Solution I
Figure S1. Sequence alignment of promoters associated with the translational machinery. Panel A shows ungapped sequences (+5 to -60) aligned to the ‘-10 box’ (consensus sequence of TAnnnT), which was identified using MEME (Bailey & Gribskov, 1998) and an initial search window of -1 to -15. Panel B as A, except gaps have been introduced 6 nt upstream of the -10 boxes to maximise alignment to a second conserved hexanucleotide sequence (GnTTnG), which was identified in the alignment shown in panel A. The second sequence is labelled as ‘-35 box’. Highlighting indicates nucleotide matches to the consensus sequences.
Figure S1. Sequence alignment of promoters associated with the translational machinery.

Panel A shows ungapped sequences (+5 to -60) aligned to the ‘-10 box’ (consensus sequence of TAnnnT), which was identified using MEME (Bailey & Gribskov, 1998) and an initial search window of -1 to -15. Panel B as A, except gaps have been introduced 6 nt upstream of the -10 boxes to maximise alignment to a second conserved hexanucleotide sequence (GnTTnG), which was identified in the alignment shown in panel A. The second sequence is labelled as ‘-35 box’.

Highlighting indicates nucleotide matches to the consensus sequences.
Figure S2. Degradation of 5'-triphosphorylated RNA using TEX. Total *P. acnes* RNA was isolated as described in Materials and Methods. *E. coli cspA* transcript was synthesised by *in vitro* transcription using T7 RNA polymerase (Invitrogen) using condition stated by the manufacturer. 0.5 μg of *cspA* was added to 1.0 μg of total RNA and treated with TEX, reactions were purified by phenol:chloroform extraction and analysed using 1.2% (w/v) agarose gel electrophoresis. Lane 1 shows the control sample before treatment. Lane 2 and 3 shows the effect without (reaction buffer only) and with TEX treatment.
7 References


