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, Faulty 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.

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

List of Figures

List of Tables

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 aicd
°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
K_2HPO_4 – di-potassium hydrogen phosphate
KEGG – Kyoto Encyclopaedia of Genes and Genomes
$KH2PO_4$ – potassium di-hydrogen phosphate
K_m – 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 pyprophosphatase
TBE – Tris-borate-EDTA
TEX – Terminator™ 5' 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 suffers 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 dehydroepiandrosterone 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 $10\text{-}10^4$ 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 acnes 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 though 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-androstanedione, 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 (doxycline, minocycline and lymecyline) (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 spironlactone, 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 pyrogenes*) 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 B_5 and Vitamin B_6 (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 lymphotoxin 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, Hunyadkürti *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).

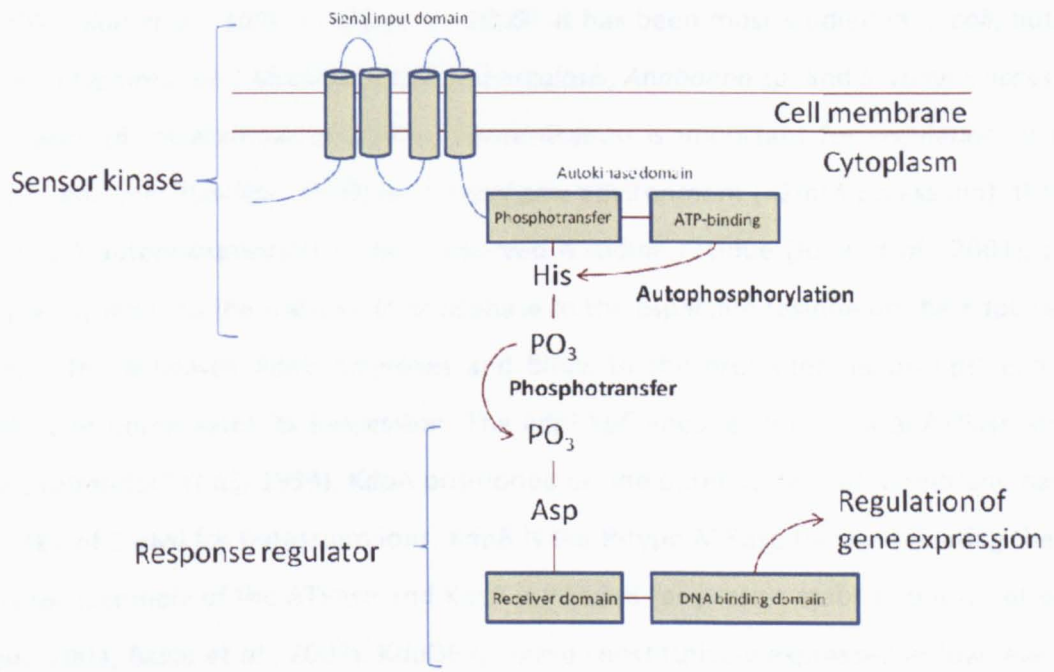


Figure 2.1 Schematic diagram of a basic two-component signal transduction system. The signal input domain is mostly, but not exclusively, associated with the cell membrane. The phosphotransfer and ATP-binding sub-domain of autokinase domain catalyse the autophosphorylation of conserved histidine residue. 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.*,

2000). 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 recruit 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 (Suelter, 1970). In a hypotonic environment (<2mM 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] N_2, 10% [v/v] CO_2, and 10% [v/v] H_2. 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 OD_{600} of 0.2. The culture was then incubated to an OD_{600} 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 OD_{600} 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 OD_{600} 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 cm^2 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 Biomedicals). Tubes were then placed in high-speed benchtop homogenizer (FastPrep®-24, MP Biomedicals; 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: $PPA0010_F$: 5'-CCCGTACTGGTCAGCGTTTA-3'; $PPA0010_R$: 5'-GCCGTCTGCTTGTACAGGTT-3'; $PPA0116_F$: 5'-CGGCAAGCAACTACTCATCA-3'; $PPA0116_R$: 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 OD_{600} 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 OD_{600}. Cells cultured in HSM had a tendency to clump and settle in the bottom of the culture flask. Therefore, for OD_{600} 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 OD_{600} of 2.0. Thus, an OD_{600} 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 $Log_{10}(OD_{600})$ 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 (KH_2PO_4) and di-potassium hydrogen phosphate (K_2HPO_4) 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*

al., 1999). 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.

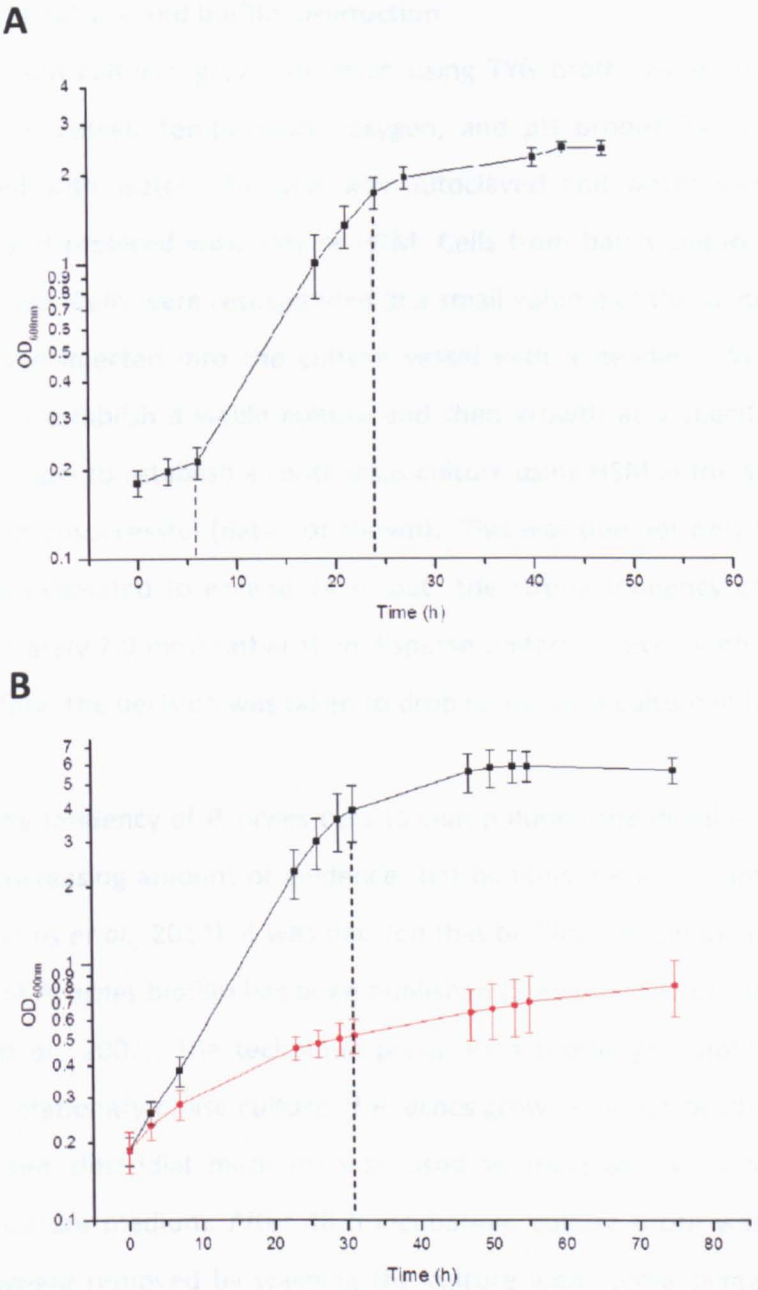


Figure 2.2 Growth curves of *P. acnes* static batch culture in different liquid media. Panel A shows the average OD_{600} of 4 independent growth curves of *P. acnes* in TYG broth. Panel B shows the average OD_{600} reading of 5 independent *P. acnes* in HSM with and without added potassium, which are 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).

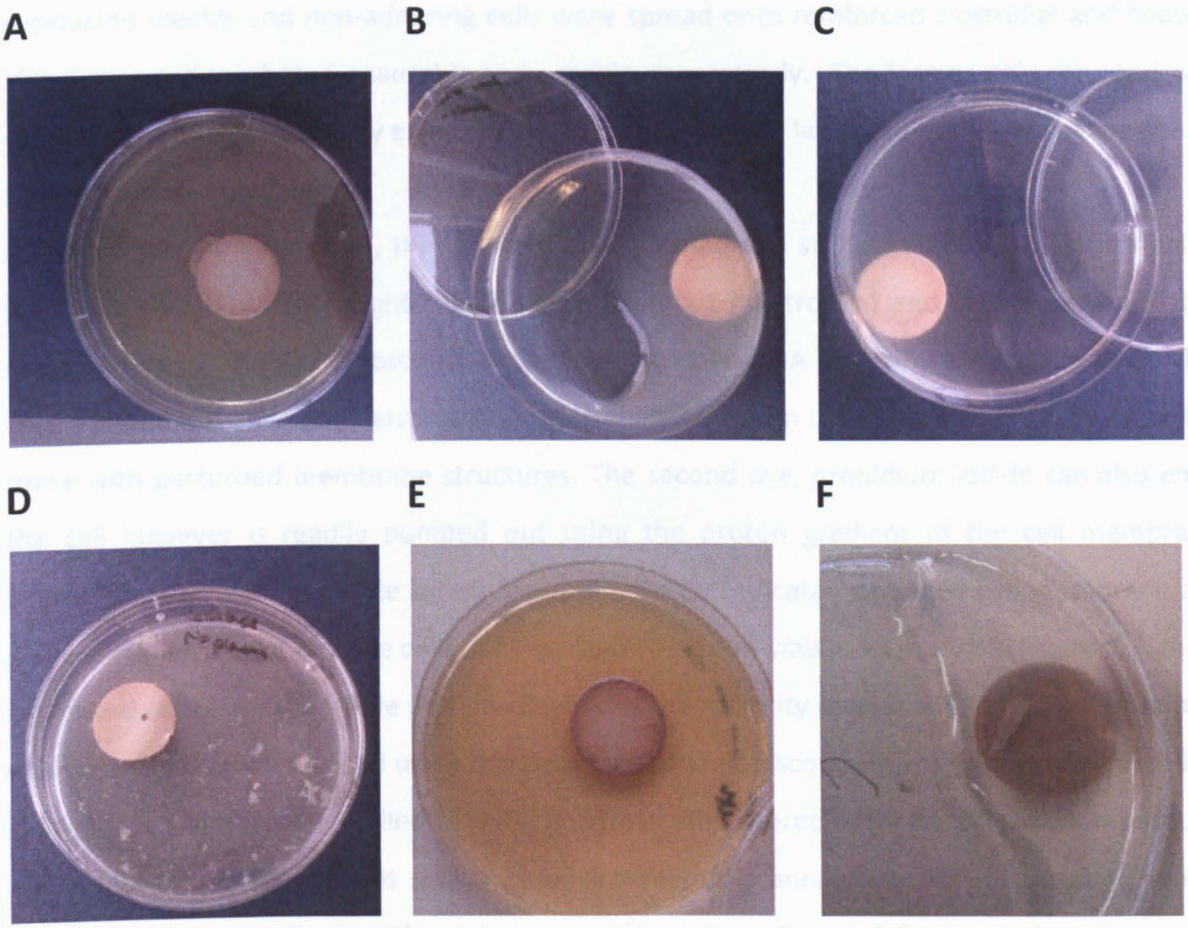


Figure 2.3 Nitrocellulose disc model of *P. acnes* biofilm culture. A nitrocellulose disc was inoculated with *P. acnes* grown in TYG broth as previously described. Panel A shows the biofilm grown on RCA after 1 week. Panel B shows the removal of non-adherent cells by saline washes. Panel C shows the 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.

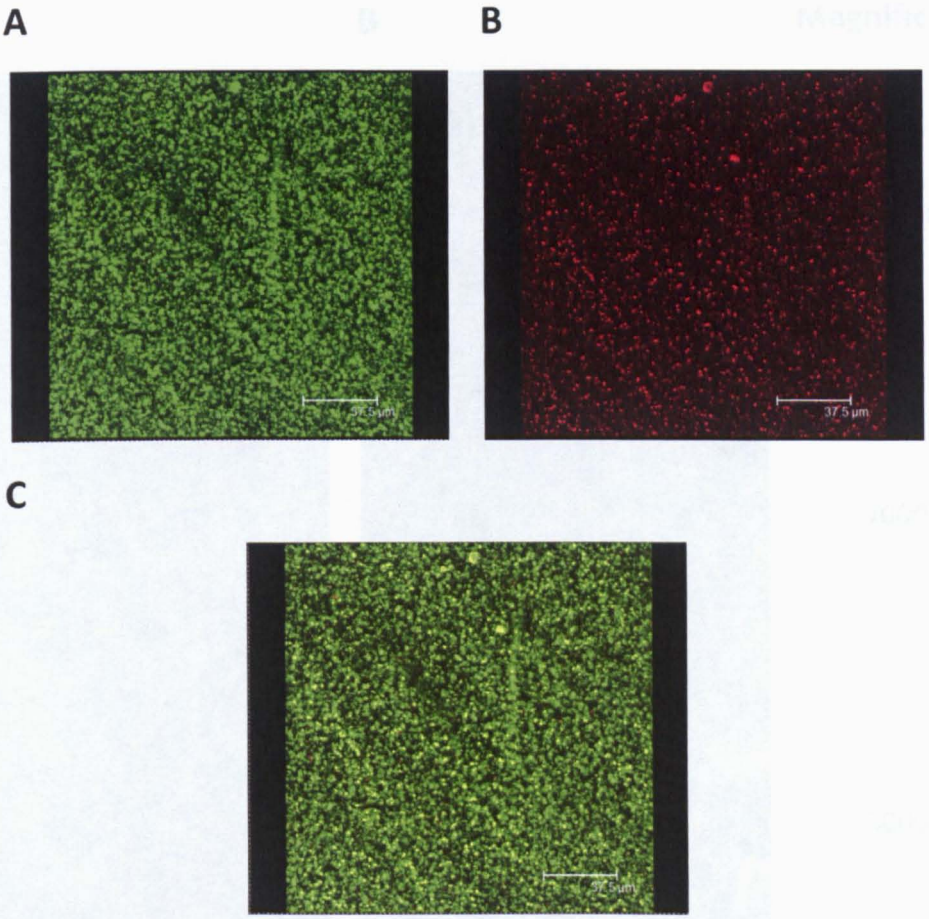
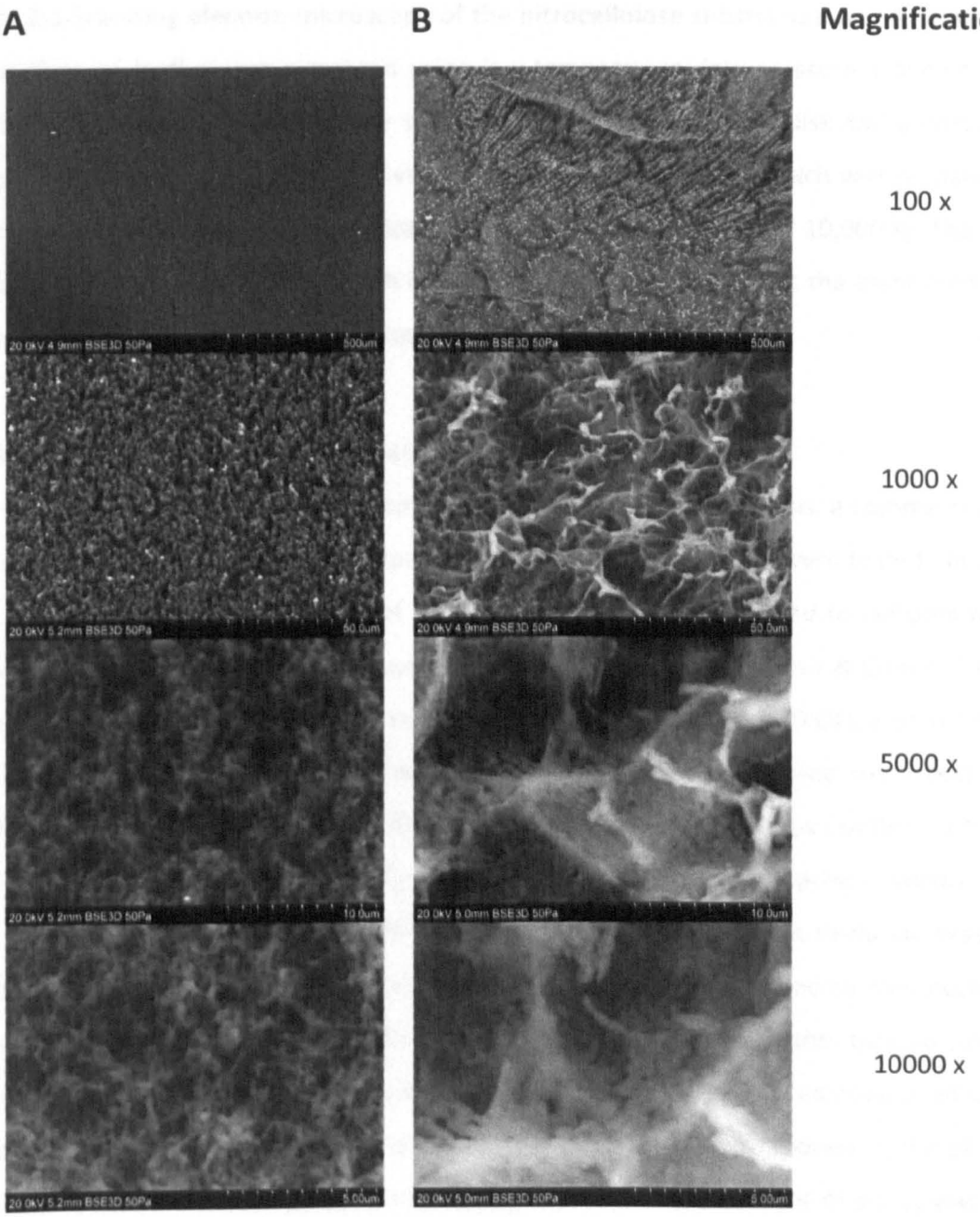


Figure 2.4 Confocal laser scanning microscopy of *P. acnes* biofilm. Biofilm that was been grown for 4 week was stained with LIVE/DEAD® BacLight™ Bacterial Viability kit (Invitrogen) to assess *P. acnes* viability using confocal laser scanning microscopy. The image was captured every 10 µm from the base 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_{600} from TYG broth. The commercial kit was RiboPure™-Bacteria (Ambion). The associated extraction method involved homogenising cells suspended in RNA_{WIZ}, 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_{600} 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).

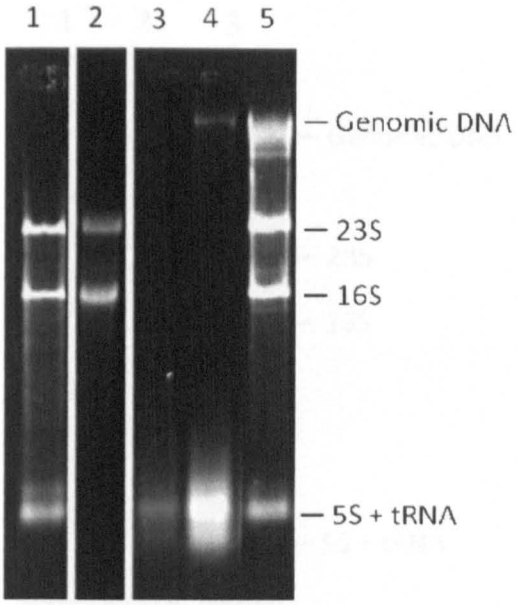


Figure 2.6 Analysis of total nucleic acid isolated using different techniques by gel electrophoresis. *P. acnes* total nucleic acid was isolated using different techniques. The samples were loaded and analysed by gel electrophoresis; 1.2% [w/v] agarose. All protocols were carried out as described in 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).

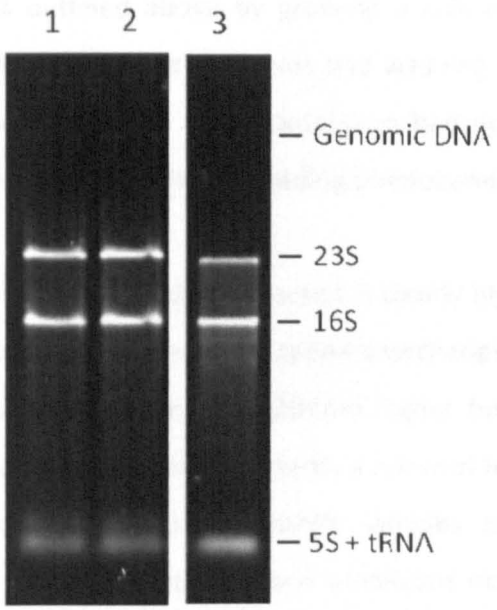


Figure 2.7 Analysis of RNA isolated from *P. acnes* grown in liquid with and without potassium downshift and as a biofilm. All samples were analysed by electrophoresis using a 1.2% [w/v] agarose gel. All lanes contained *P. acnes* total RNA. Lane 1 and 2 show total RNA isolated from cells without downshift and with downshift, respectively. Lane 3 shows the total RNA isolation from a 4 week-old biofilm. For details of isolation procedures see text.

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 *kdpB* transcript is clearly higher following the potassium downshift, whilst the level of the *gyrA* transcript appears unchanged. Densitometric analysis of the gel revealed that the *kdpA* amplicon was 220-fold higher following potassium downshift (values normalised to *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 *kdpB* and *gyrA* amplicons migrated as expected for their predicted size (179 and 85 bp, respectively).

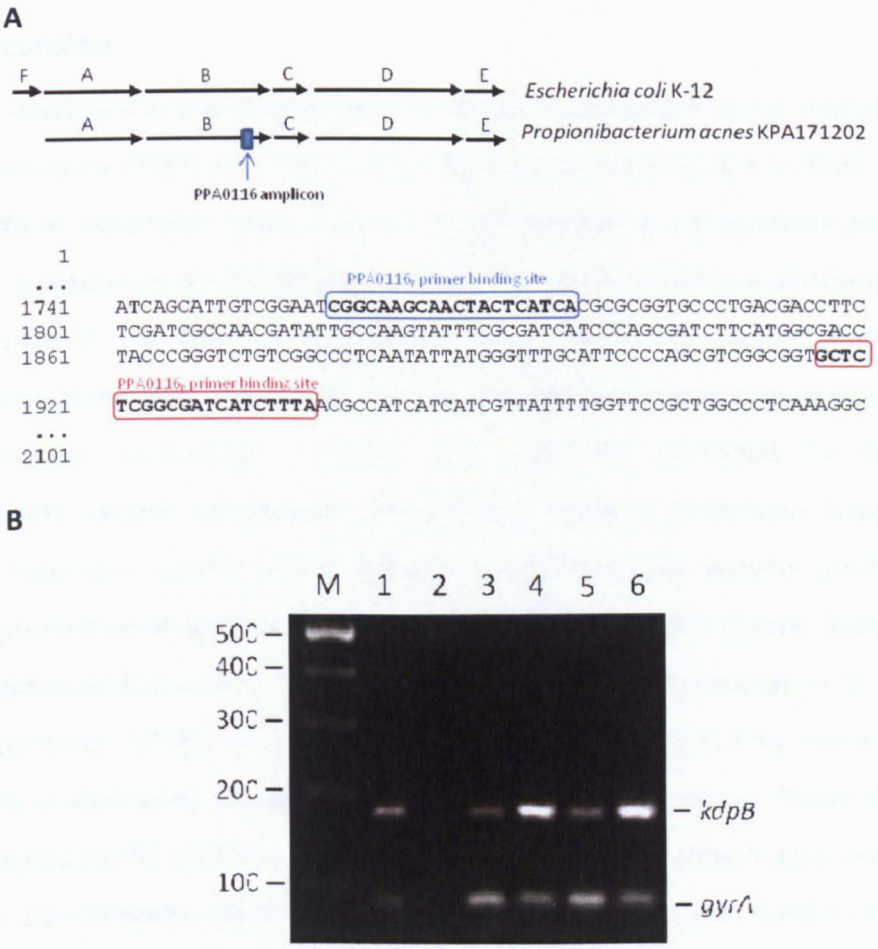


Figure 2.8 Analysis of RT-PCR product from RNA isolated with and without potassium downshift. Complementary DNA was synthesised from equal amount of *P. acnes* RNA with and without downshift, and used as template for PCR amplification of segments of the target genes PPA0010 (*gyrA*) and PPA0116 (*kdpB*) 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 (Coenye *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. acnes*, 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 TI267747_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 RankProdIt (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 *MICROBE*xpress™-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 (μ) and standard deviation (σ) 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).

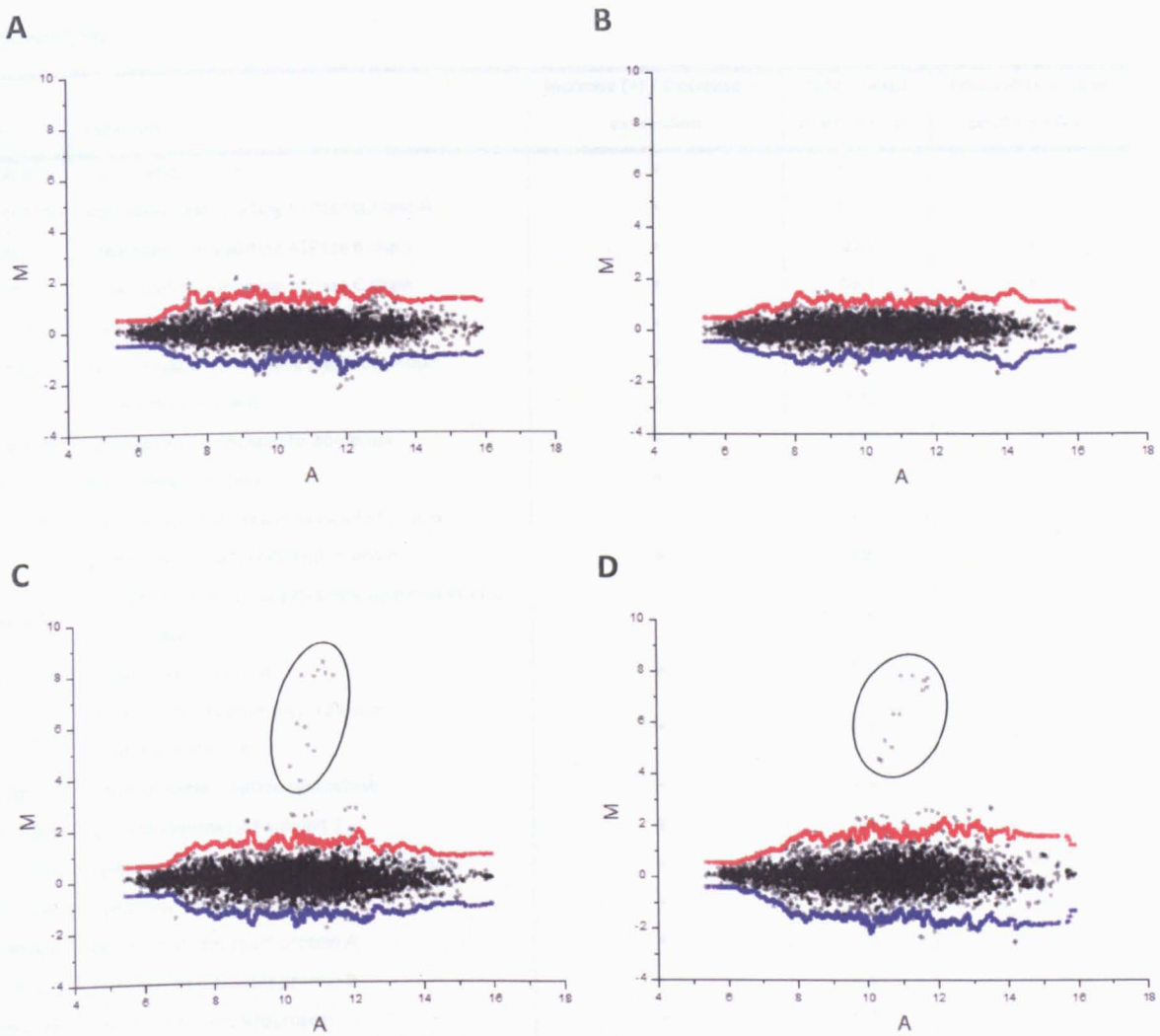


Figure 3.1 M-A scatterplot of the gene expression values from microarray analysis. In panel A, the M values equal $\log_2$ (without-downshift sample 2/ without -downshift sample 1), and A equals ($\log_2$ without -downshift sample 2+ $\log_2$ without -downshift sample 1)/2. Panel B as A, but substituted 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 (μ) and standard deviation (σ) 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.

Gene	Function	Increase (+) / Decrease (-) expression	Fold change in expression	Probability of false positive < 0.15
PPA0114	hypothetical protein	+	201.4	+
PPA0115	potassium-transporting ATPase subunit A	+	242.3	+
PPA0116	potassium-transporting ATPase B chain	+	224	+
PPA0117	potassium-transporting ATPase C chain	+	69.1	+
PPA0118	two-component sensor, KdpD	+	19.5	+
PPA0119	two-component response regulator, KdpE	+	33.9	+
PPA0120	hypothetical protein	+	5.9	+
PPA0476	glucosamine-6-phosphate isomerase	+	4.6	-
PPA0667	hypothetical protein	+	2.5	-
PPA1091	hypothetical membrane associated protein	+	3.1	-
PPA1092	protein with Sua5/YciO/YrdC domain	+	2.8	-
PPA1093	5`-methylthioadenosine/S-adenosylhomocysteine nuclosidase	+	2.4	-
PPA1105	phosphoglucomutase	+	2.1	+
PPA1224	putative glycerophosphoryl diester phosphodiesterase	+	2.9	-
PPA1287	non-ribosomal peptide synthetase	+	2.9	-
PPA1288	surfactin synthetase subunit 1	+	3.4	-
PPA1289	cysteine synthase/ornithine cyclodeaminase	+	3.6	+
PPA1290	cystathionine beta-synthase	+	4.1	+
PPA1676	ferrous iron transport protein A	+	3.8	-
PPA1677	ferrous iron transport protein B	+	3.8	-
PPA1758	outer membrane lipoprotein	+	5.9	+
PPA1759	ABC transporter ATP-binding protein	+	5.0	+
PPA1760	ABC transporter associated permease	+	3.8	-
PPA2286	phosphoglucomutase/phosphomannomutase	+	2.9	-
PPA2268	alanine dehydrogenase	+	1.6	+
PPA0012	L-lactate dehydrogenase	-	2.3	-
PPA0557	sodium-and chloride-dependent transporter	-	3.8	-
PPA0964	transcriptional regulator	-	2.9	-
PPA1476	glycine betaine transport system permease protein	-	2.1	-
PPA1807	hypothetical protein	-	4.2	-
PPA2152	putative peptide transport system secreted peptide-binding protein	-	3.0	-
PPA2175	rare lipoprotein A (RlpA) family protein	-	4.1	+

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 K_2HPO_4 and KH_2PO_4 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.

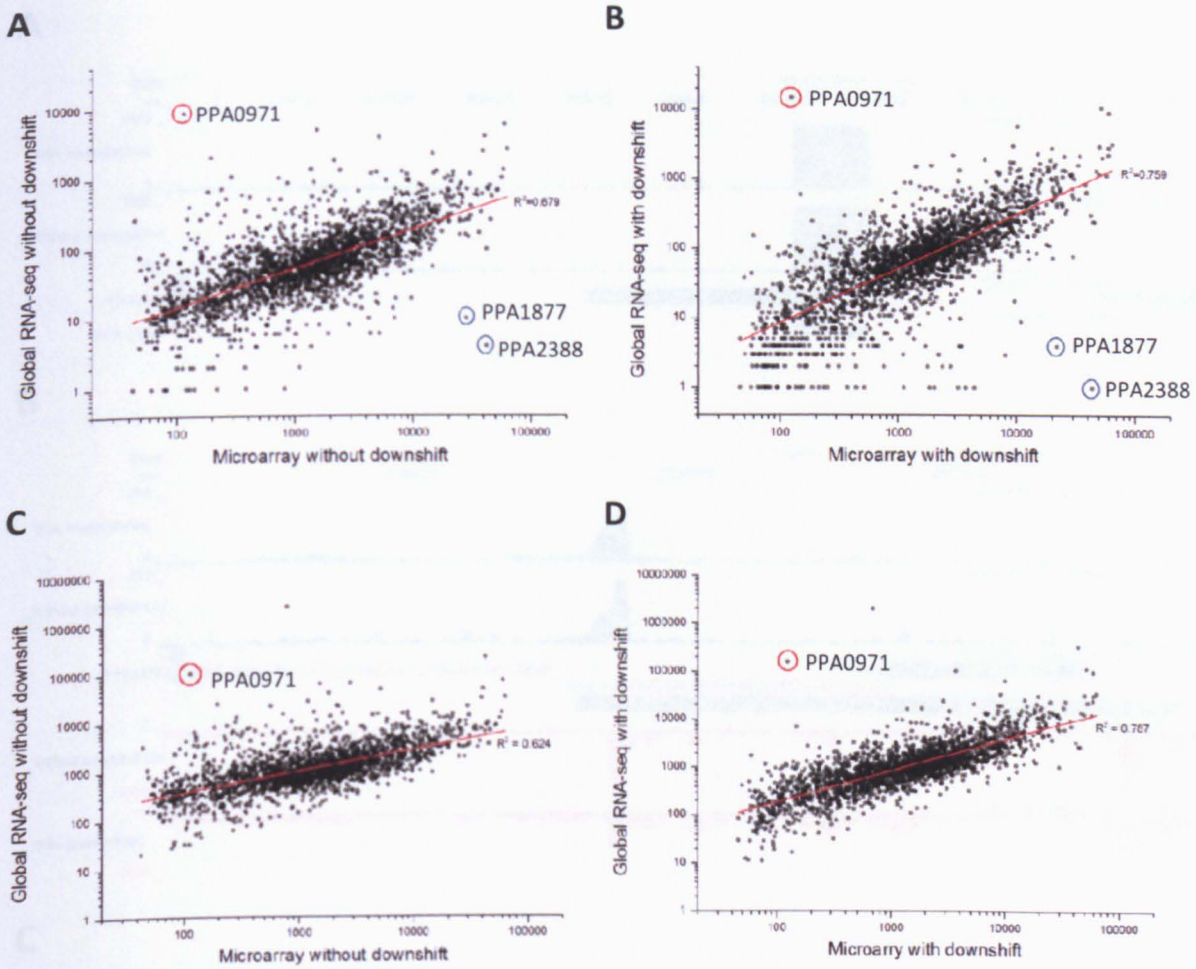
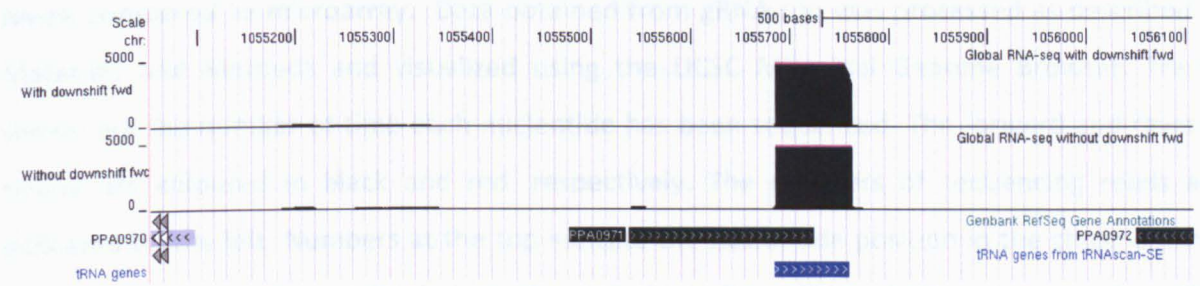
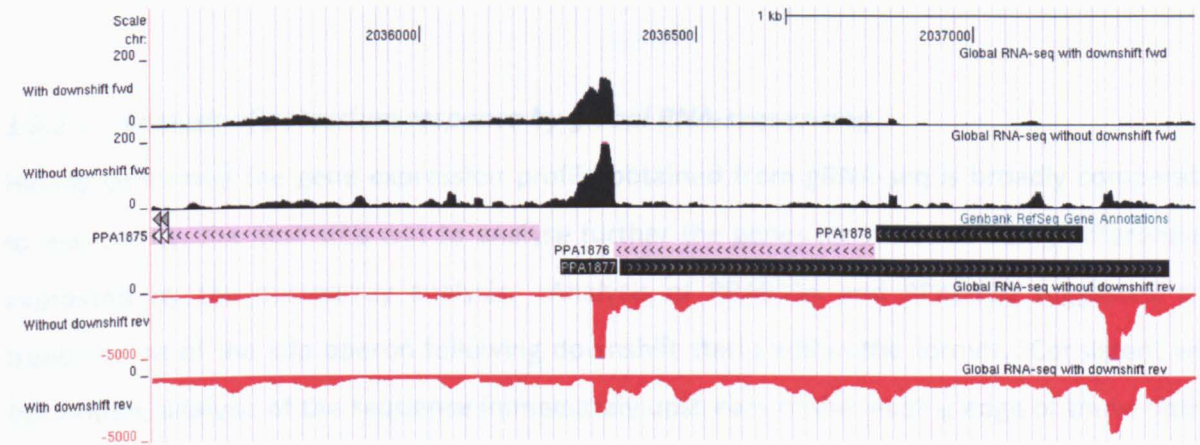


Figure 3.2. Scatter plot of average expression of protein coding genes from microarray and global RNA-sequencing. Start and end positions of the protein coding regions for *P. acnes* were obtained from NCBI database. The Average expression per gene was obtained by dividing the sum of the sequencing values, 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.

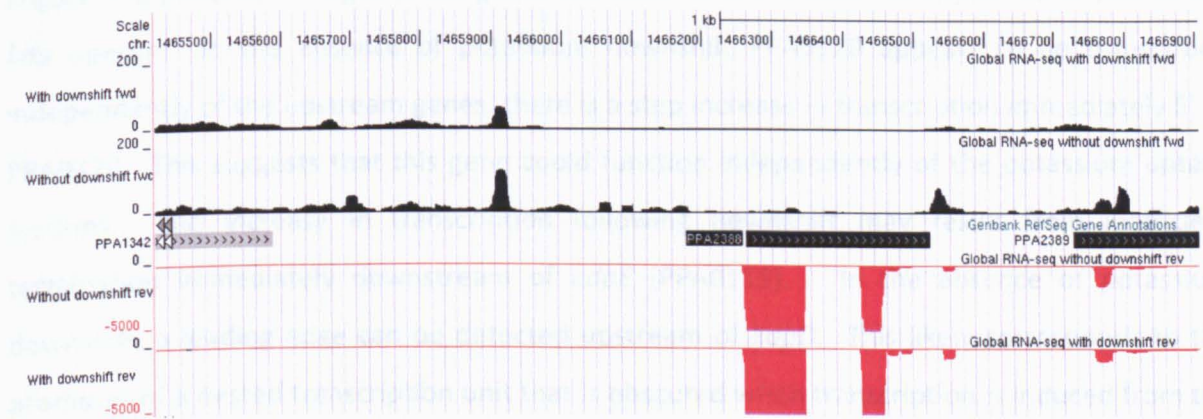
A



B



C



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

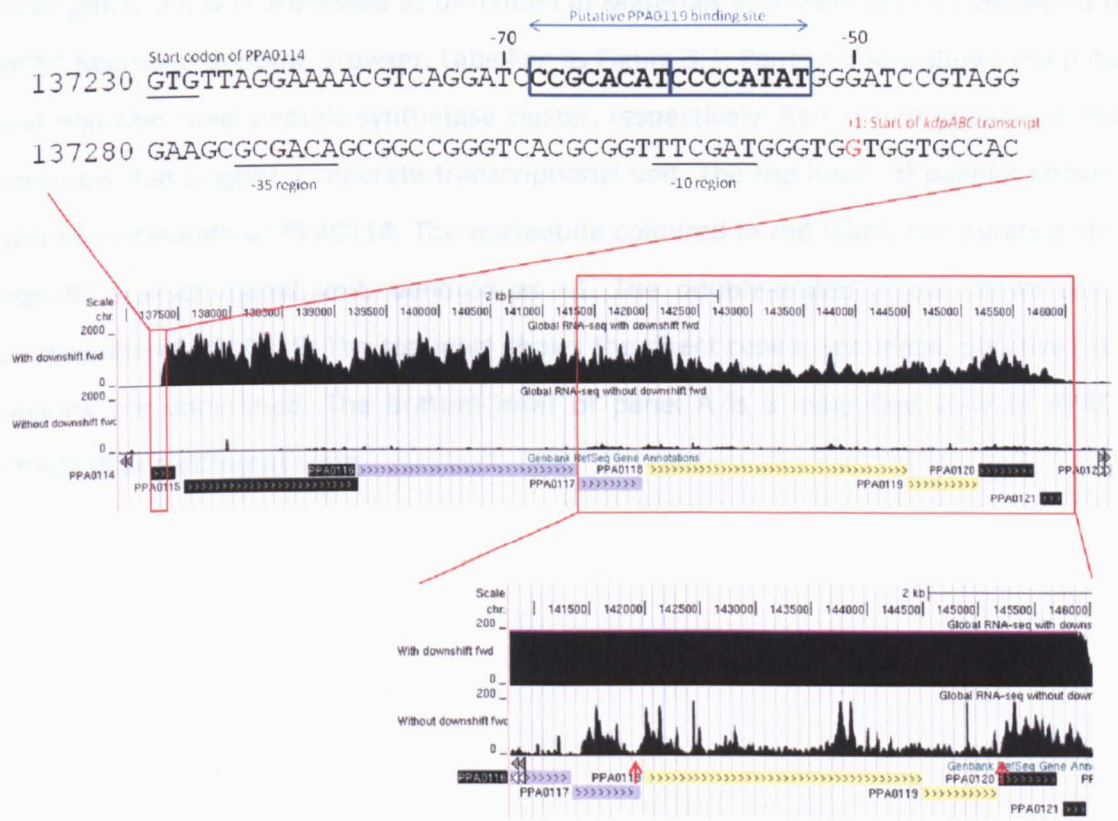
Having confirmed the gene expression profile obtained from gRNA-seq is broadly comparable to microarray, the next step was to analyse further the genes identified as being differentially expressed by the microarray analysis. Analysis of PPA0114 and PPA0120 suggested that transcription of the *kdp* operon 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

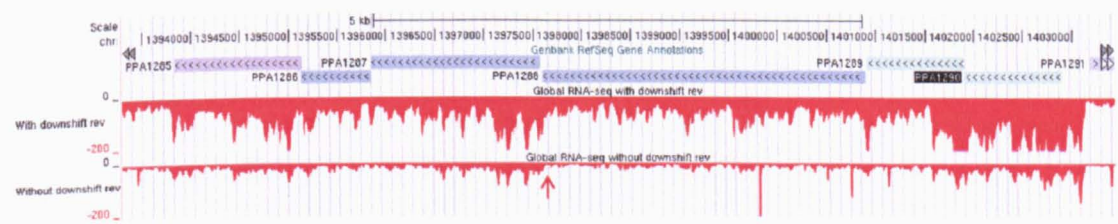
transcriptional regulation of *kdpDE* (Sugiura *et al.*, 1993, Sugiura *et al.*, 1992). In the presence of potassium, *kdpDE* was found to be transcribed as a transcriptional unit independent of *kdpABC* in *E. coli*. The transcriptional start site for the putative *kdpD* transcript in *P. acnes* correlated 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.

A



B



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.

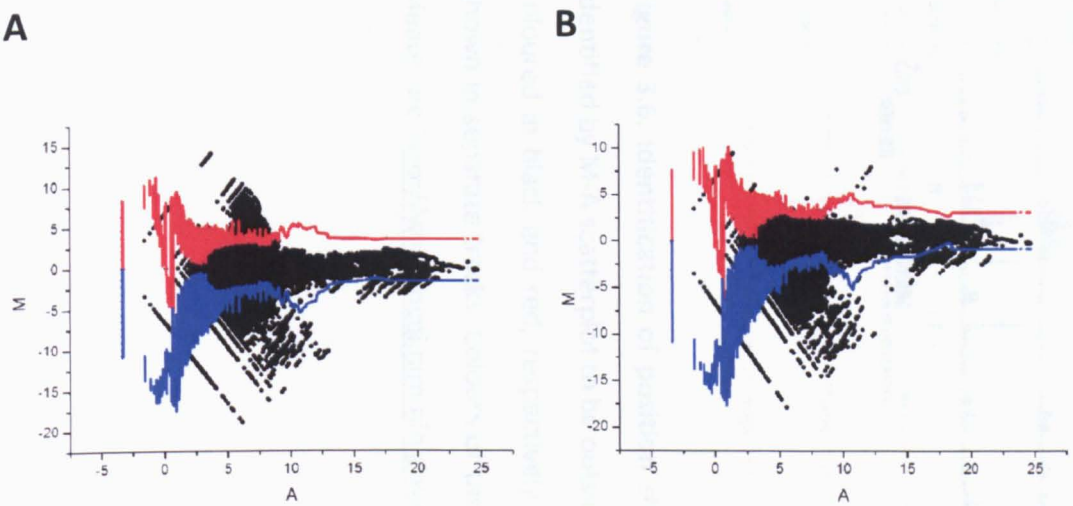


Figure 3.5 M-A scatter plot of global RNA-sequencing reads. Generation of A and M values and the 'noise' envelope (red and blue trace) were as described in Figure 3.1, except using the sequencing reads. The average and standard deviation were calculated from a sliding window of the 5000 corresponding 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.

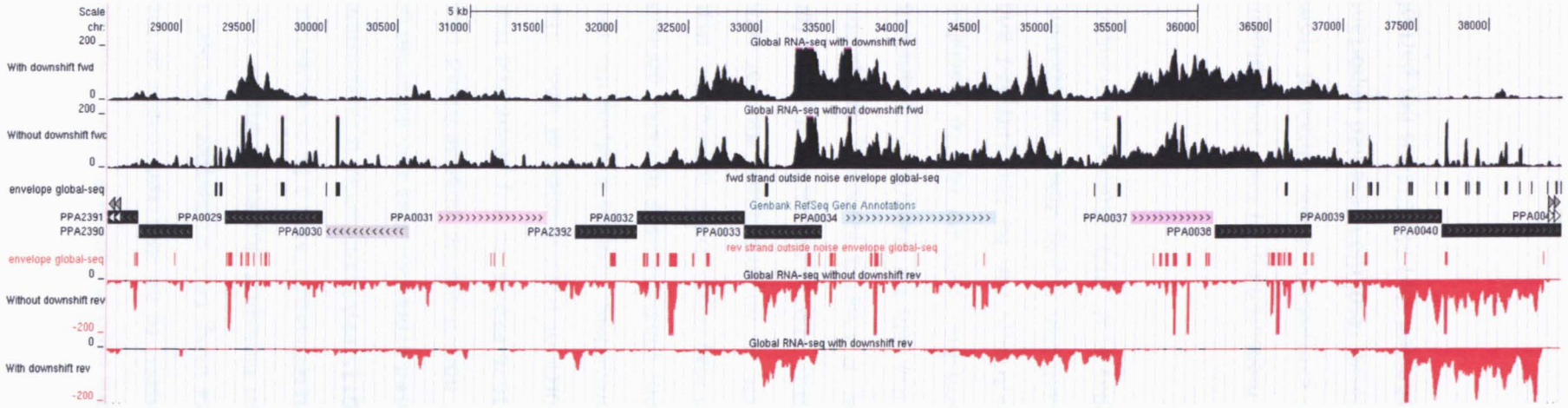


Figure 3.6. Identification of position showing differential expression by gRNA-seq in *P. acnes*. Track shows the position identified by M-A scatterplot to be outside the defined noise envelop. Labelling as Figure 3.3. Forward and reverse strand are coloured in black and red, respectively. For each strand, 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/

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_2HPO_4 and KH_2PO_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 *MICROBE*xpress™-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 (μ) and standard deviation (σ) 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 S1).

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.*,

2007), (Table S3) that could account for the large number of PDSs detected in the transcriptome.

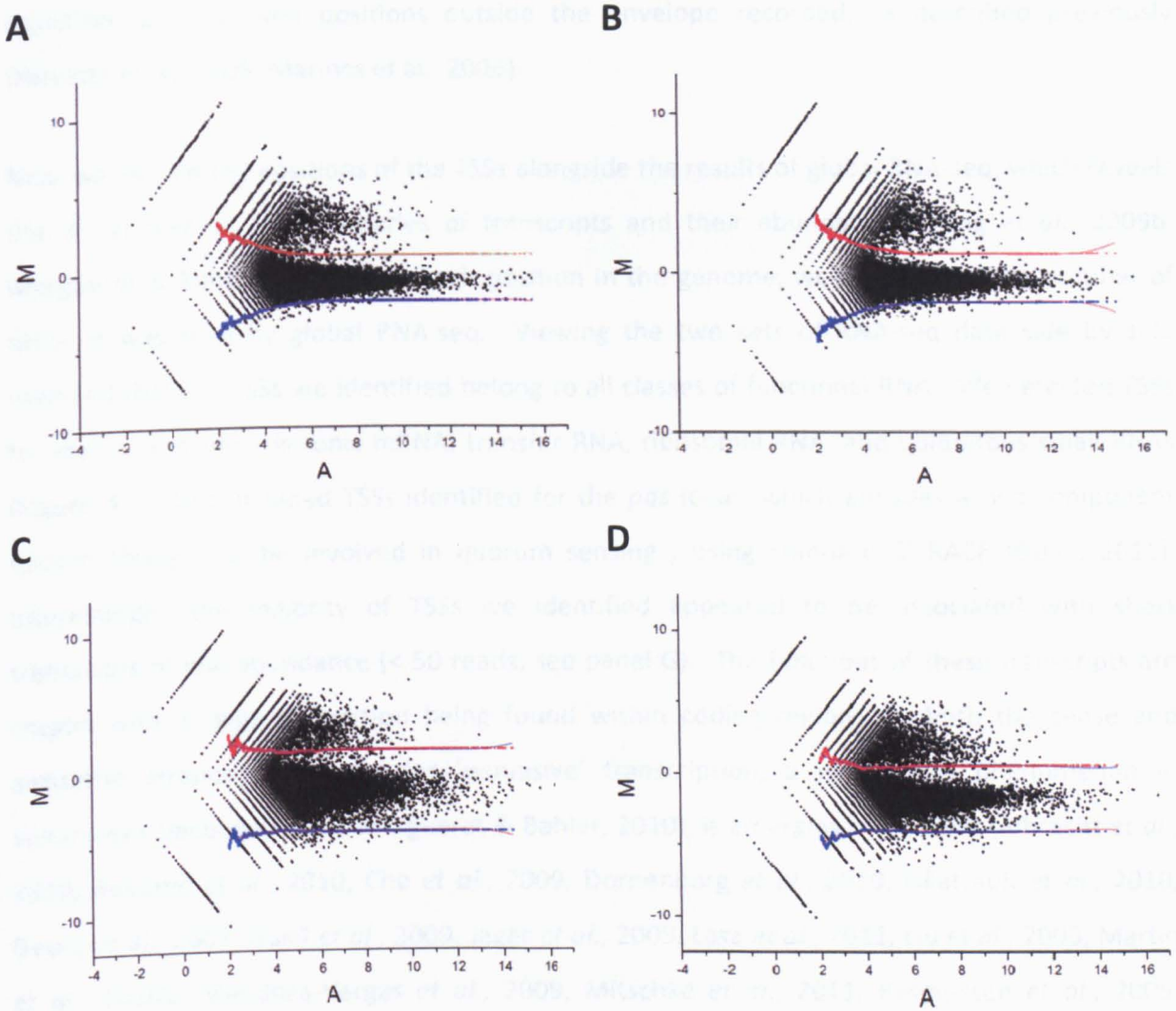
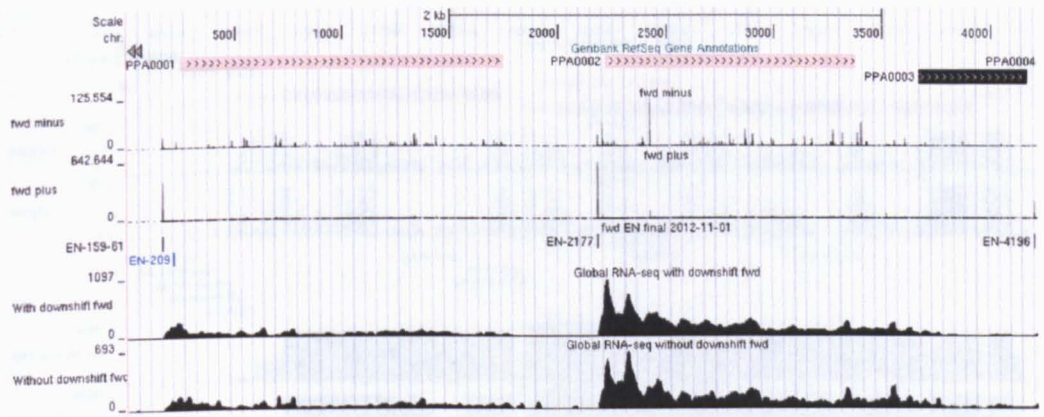


Figure 4.1. M-A scatterplots of values from differential RNA-seq. Panels A and B show represent the data for cells cultured without and with a potassium downshift, respectively. Panels C and D, as A and B, except data is for a duplicate pair of cultures. The M values correspond to Log_2 (Plus/Minus) and A values to (Log_2 Plus + Log_2 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

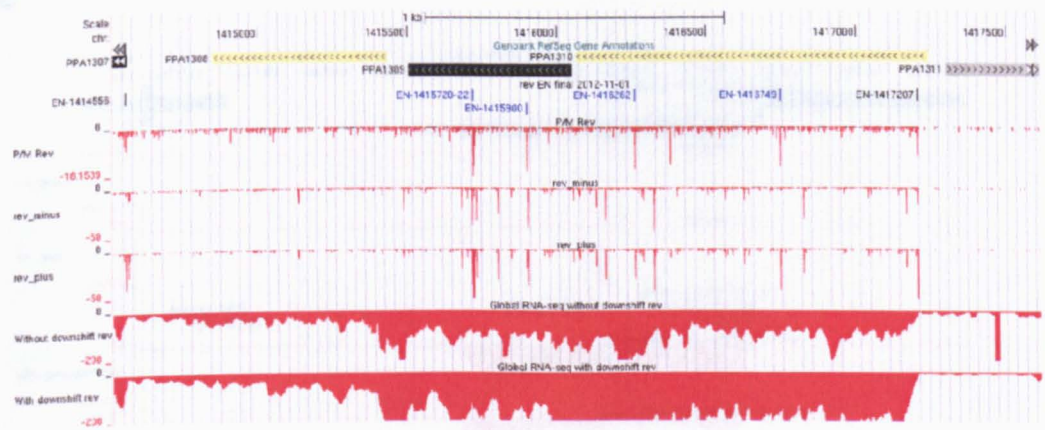
the envelope containing sites of processing. The boundaries were defined by sorting the pairs of values in ascending order of A, and then calculating the average (μ) and standard deviation (σ) of M in a moving window of 5,000 pairs. Upper and lower envelopes were defined by the equation: $\mu \pm 3\sigma$, 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).

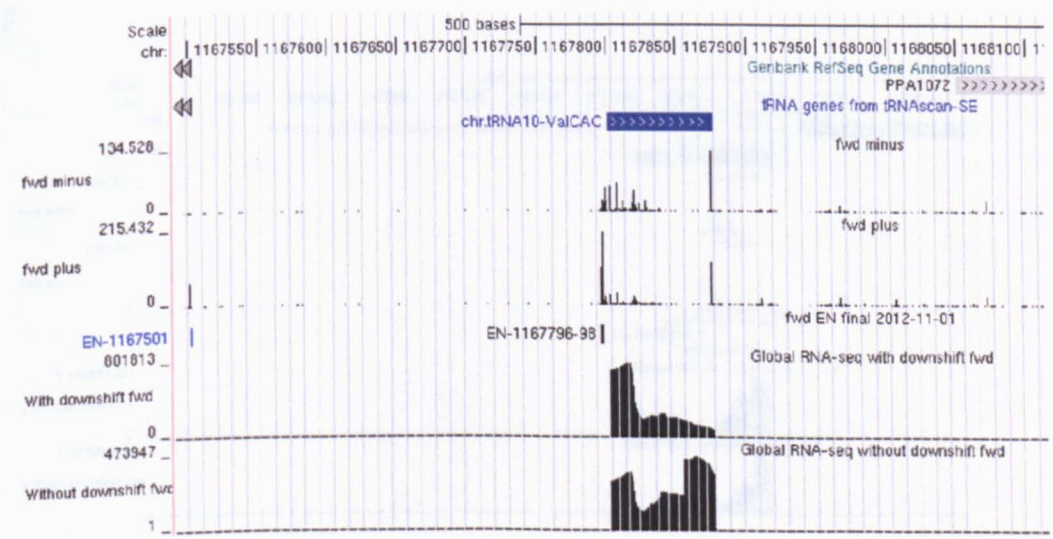
A


Scale
chr:
2 kb
500
1000
1500
2000
2500
3000
3500
4000
Genbank RefSeq Gene Annotations
PPA0001
PPA0002
PPA0003
PPA0004
fwd minus
125.554
fwd minus
0
fwd plus
642.644
fwd plus
0
fwd EN final 2012-11-01
EN-159-61
EN-209
EN-2177
EN-4196
Global RNA-seq with downshift fwd
1097
With downshift fwd
0
Global RNA-seq without downshift fwd
693
Without downshift fwd
0

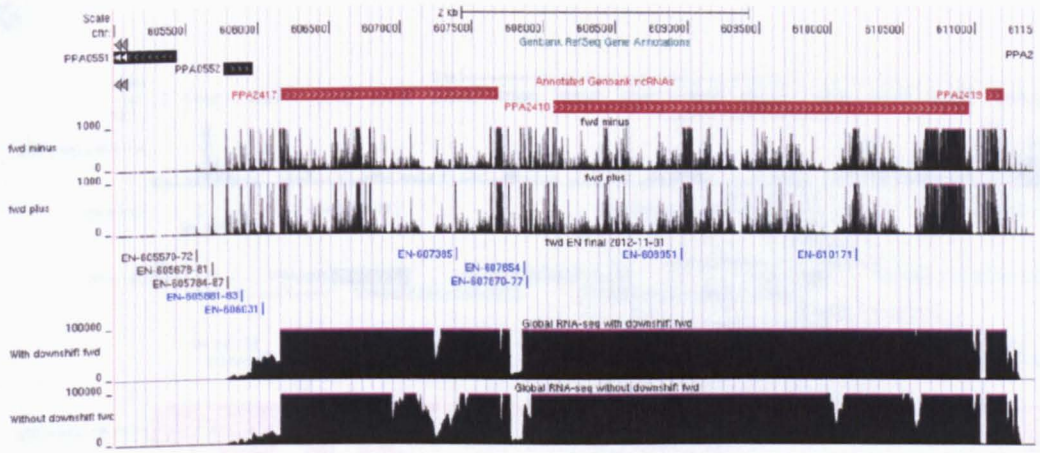
B


1 kb
Genbank RefSeq Gene Annotations
PPA1307
PPA1308
PPA1310
PPA1305
PPA1311
rev EN final 2012-11-01
EN-1414555
EN-1415720-22
EN-1415262
EN-1413749
EN-1417207
EN-1415980
P/V Rev
P/V Rev
-161539
rev_minus
rev_minus
-58
rev_plus
rev_plus
-58
Global RNA-seq without downshift rev
Without downshift rev
-230
Global RNA-seq with downshift rev
With downshift rev
-230

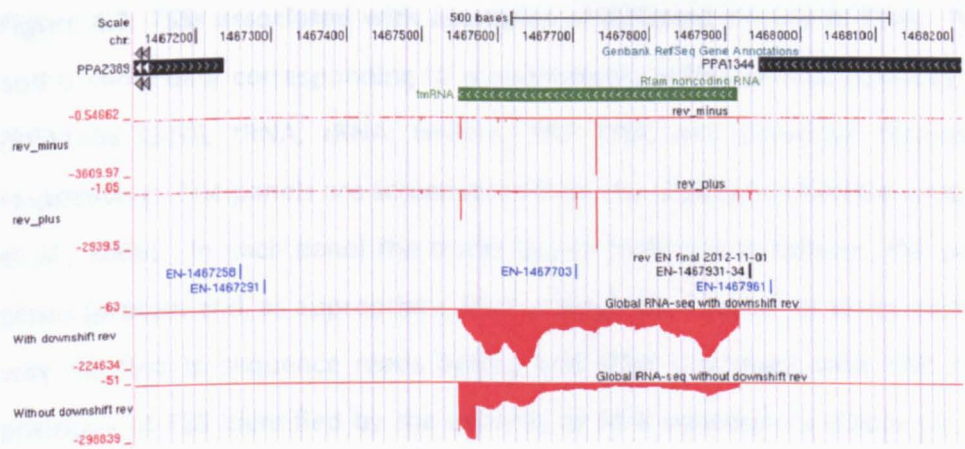
C


Scale
chr:
500 bases
1167550
1167600
1167650
1167700
1167750
1167800
1167850
1167900
1167950
1168000
1168050
1168100
Genbank RefSeq Gene Annotations
PPA1072
tRNA genes from tRNAscan-SE
chr.tRNA10-ValCAC
fwd minus
134.528
fwd minus
0
fwd plus
215.432
fwd plus
0
fwd EN final 2012-11-01
EN-1167501
EN-1167796-98
601813
Global RNA-seq with downshift fwd
With downshift fwd
0
473947
Global RNA-seq without downshift fwd
Without downshift fwd
1

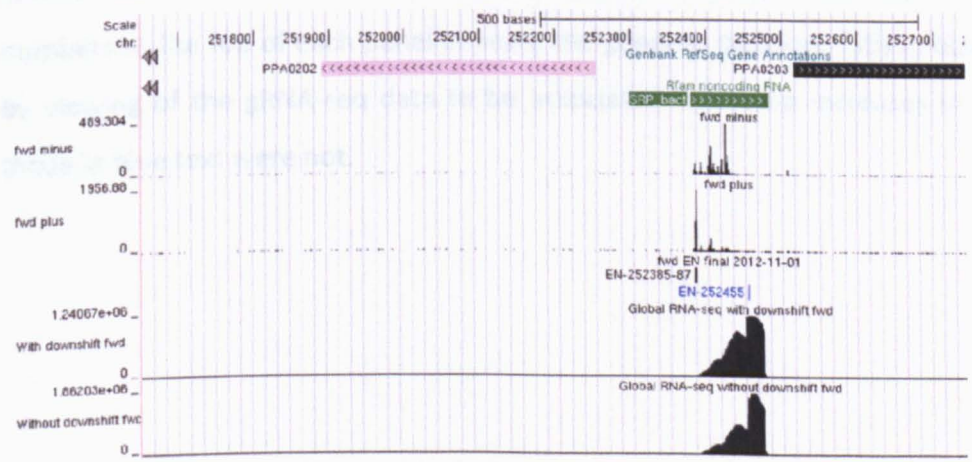
D


Scale
chr
2 kb
GenBank RefSeq Gene Annotations
PPA0551
PPA0552
Annotated Genbank ncRNAs
PPA2417
PPA2418
fwd minus
fwd plus
fwd EN final 2012-11-01
EN-606570-72
EN-606678-81
EN-606784-87
EN-606881-83
EN-606931
EN-607385
EN-607654
EN-607670-77
EN-608951
EN-610171
Global RNA-seq with downshift fwd
With downshift fwd
Global RNA-seq without downshift fwd
Without downshift fwd

E


Scale
chr
500 bases
Genbank RefSeq Gene Annotations
PPA2389
PPA1344
Rfam noncoding RNA
tmRNA
rev_minus
rev_plus
rev EN final 2012-11-01
EN-1467258
EN-1467291
EN-1467703
EN-1467931-34
EN-1467961
Global RNA-seq with downshift rev
With downshift rev
Global RNA-seq without downshift rev
Without downshift rev

F


Scale
chr
500 bases
Genbank RefSeq Gene Annotations
PPA0202
PPA0203
Rfam noncoding RNA
SRP_bact
fwd minus
fwd plus
fwd EN final 2012-11-01
EN-252385-87
EN-252455
Global RNA-seq with downshift fwd
With downshift fwd
Global RNA-seq without downshift fwd
Without downshift fwd

G

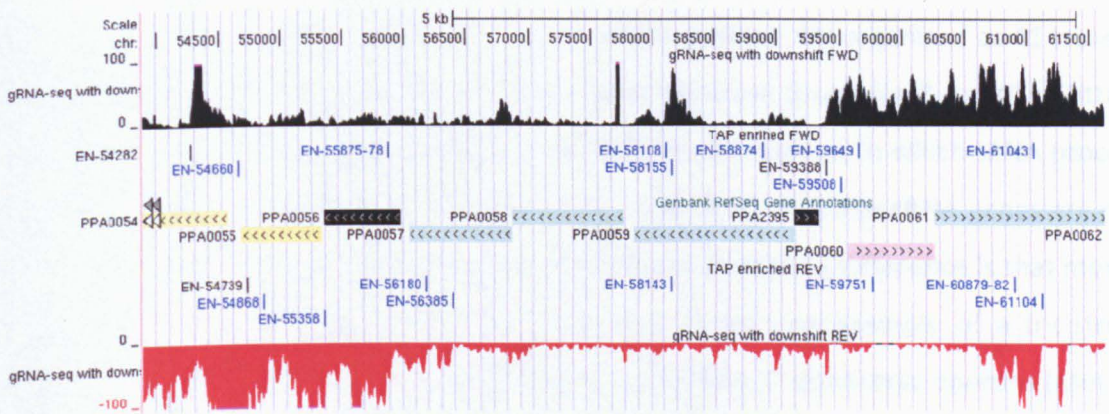


Figure 4.2. TSSs associated with examples of different classes of RNA. Panels A, B, C, D, E, F and G show data corresponding to monocistronic mRNA (PPA0001/0002), polycistronic mRNA (PPA1308-1310), tRNA, rRNA, tmRNA, SRP RNA, and pervasive transcriptional start sites, respectively. The panels are screenshots 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*.

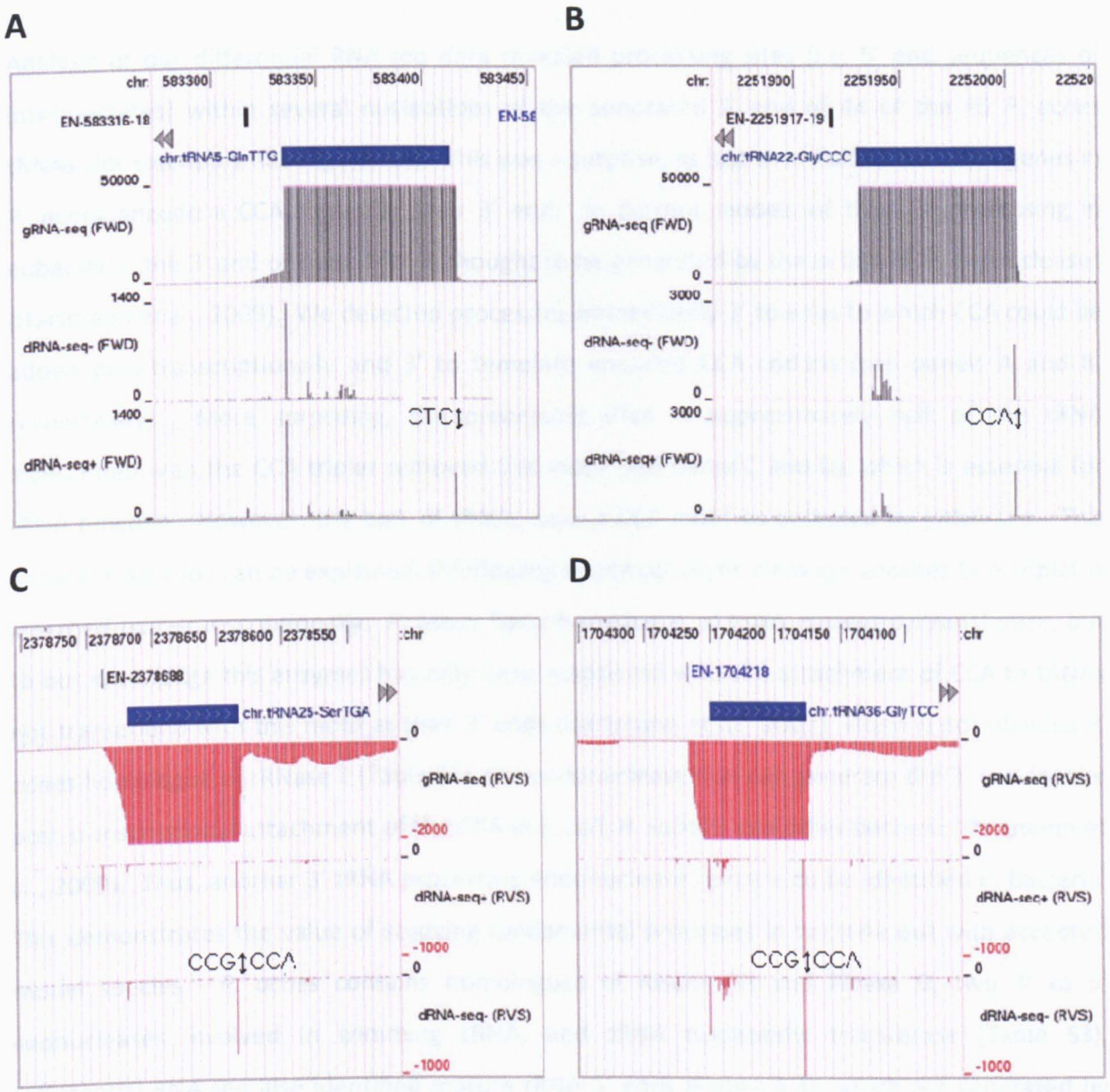


Figure 4.3. Location of 3' processing sites for tRNA. Panels A, B, C and D show correspond to tRNAs Gln, TTG (PPA2415), Gly, CCC (PPA2455), Ser, TGA (PPA2461) and Gly, TCC (PPA2438), respectively. For each panel, the tracks depict from top to bottom the position of annotated genes, the average of the 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.

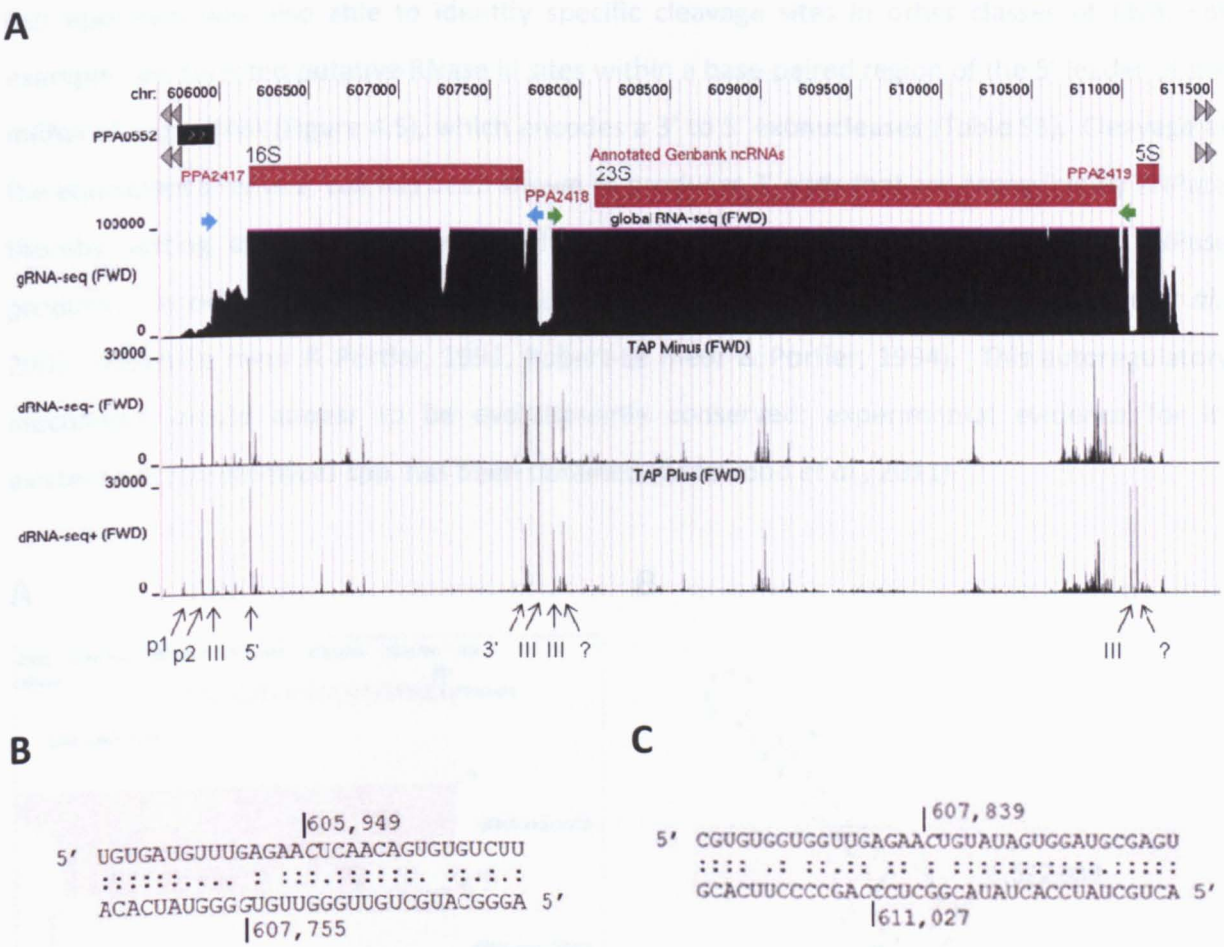


Figure 4.4. Location of ribosomal RNA processing sites. Panel A shows an annotated view of the *rrnA* cluster. The vertical arrows at the bottom indicate the position of transcriptional start sites and major processing sites referred to in the text, respectively. Predicted RNase III sites are labelled. 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).

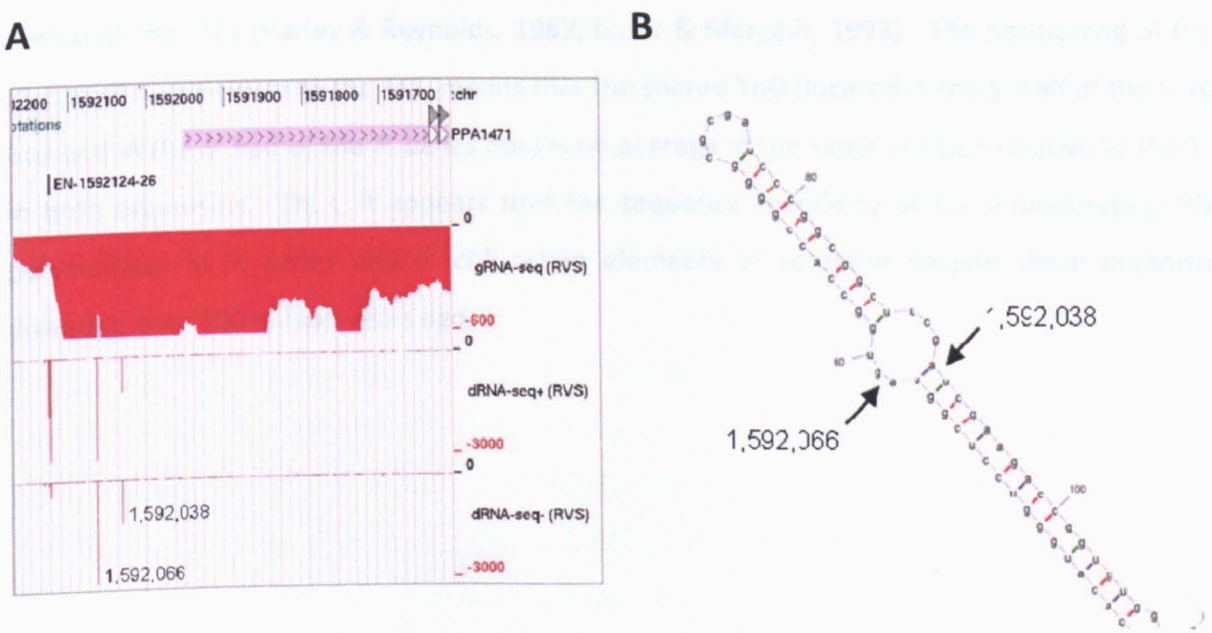


Figure 4.5. Location of RNA processing sites in 5' UTR of *pnp* mRNA. Panel A shows the RNA-seq data, while panel B shows the location of the processing sites relative to the secondary structure of the 5' UTR as predicted using Mfold (Zuker, 2003). The labelling of panel A is as Figure 4.2. Panel B, 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 GnTTnG 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.

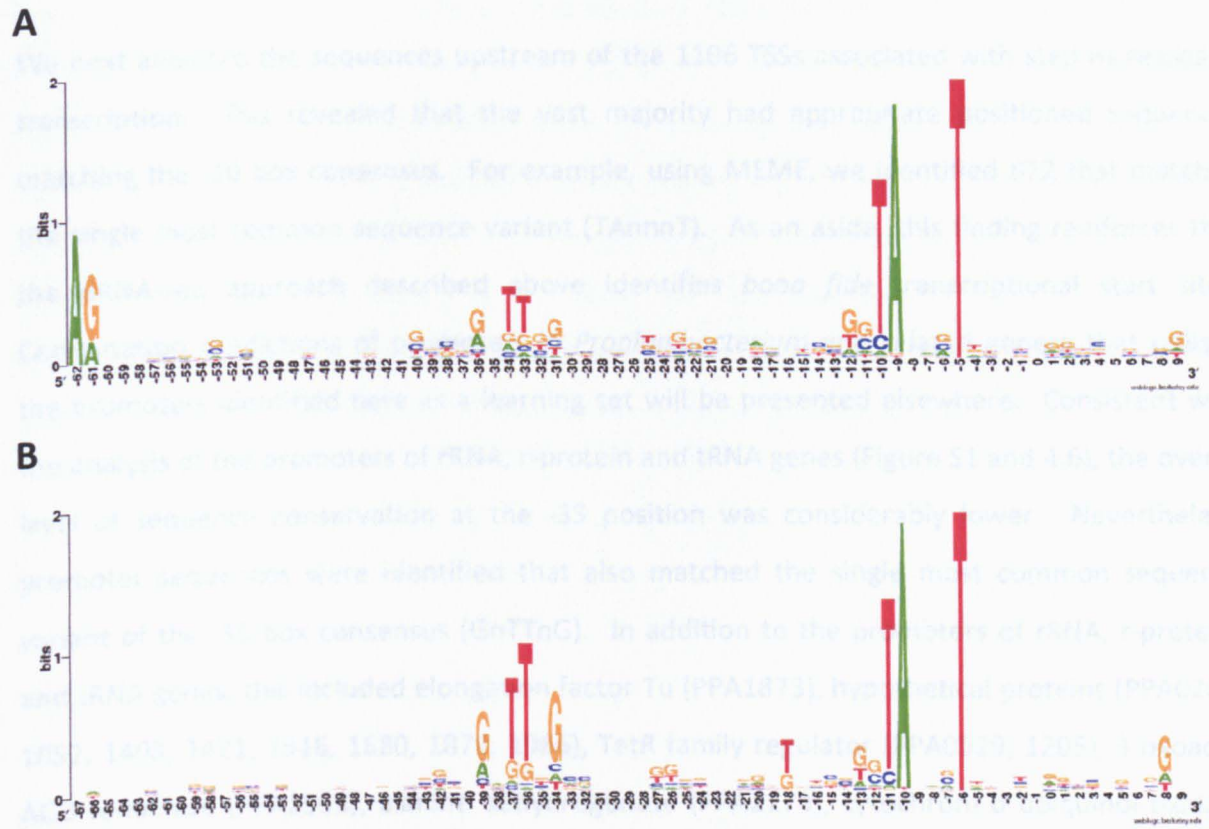


Figure 4.6. The conserved sequences of promoters associated with the translational machinery. Panels A and B are Weblogo representations (Crooks *et al.*, 2004) without and with changing the length of the spacer of individual promoters to maximise the alignment of the -35 box (Figure S1). The combined 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.

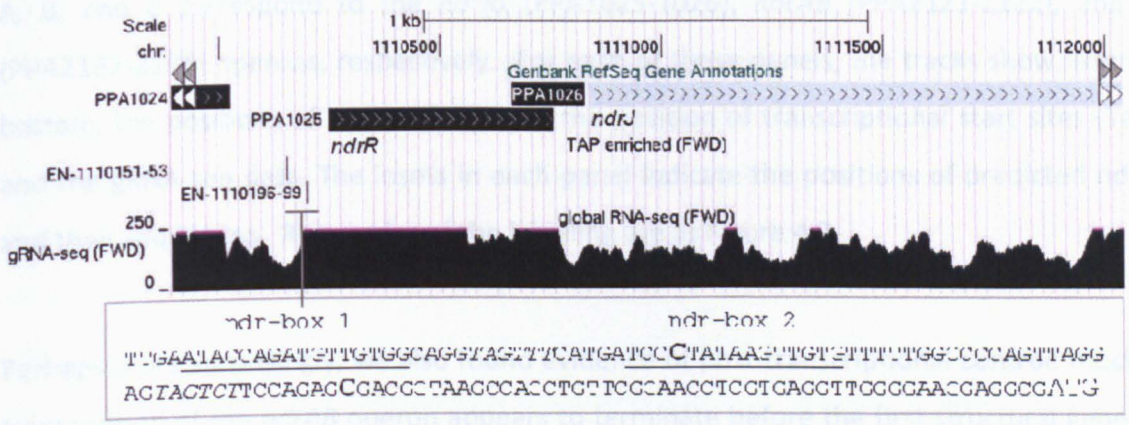
We next analysed the sequences upstream of the 1106 TSSs associated with step increases in transcription. This revealed that the vast majority had appropriate positioned sequences matching the -10 box consensus. For example, using MEME, we identified 872 that matched the single most common sequence variant (TAnnnT). 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 (GnTTnG). 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

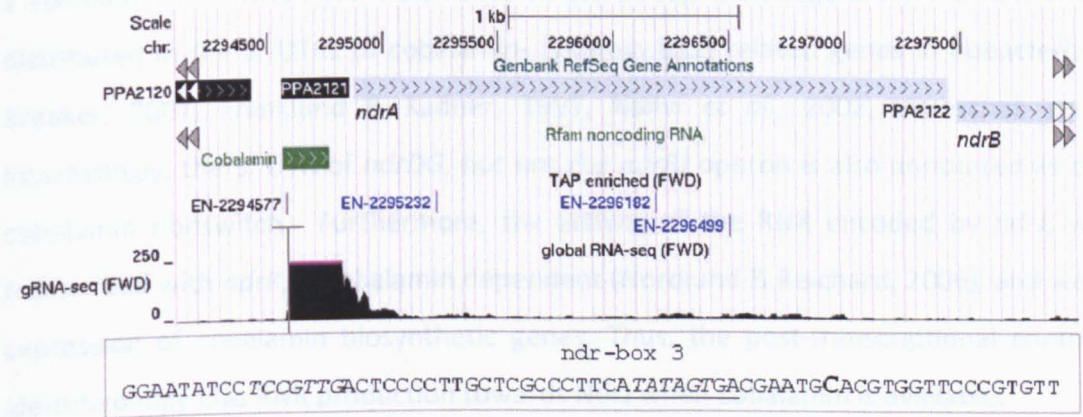
The identification of TSSs and promoter sequences alongside our high-resolution transcriptome maps provides a much improved platform for assessing the complexity of gene regulation. This is illustrated here using the *P. acnes* homologue of NdrR, a transcription factor that controls the expression of 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.

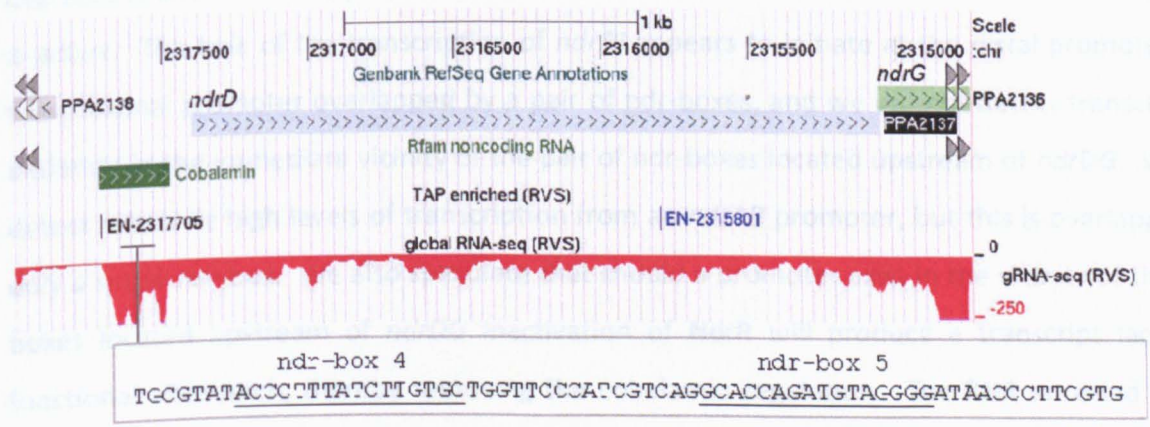
A



B



C



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 *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 *pqs* locus.

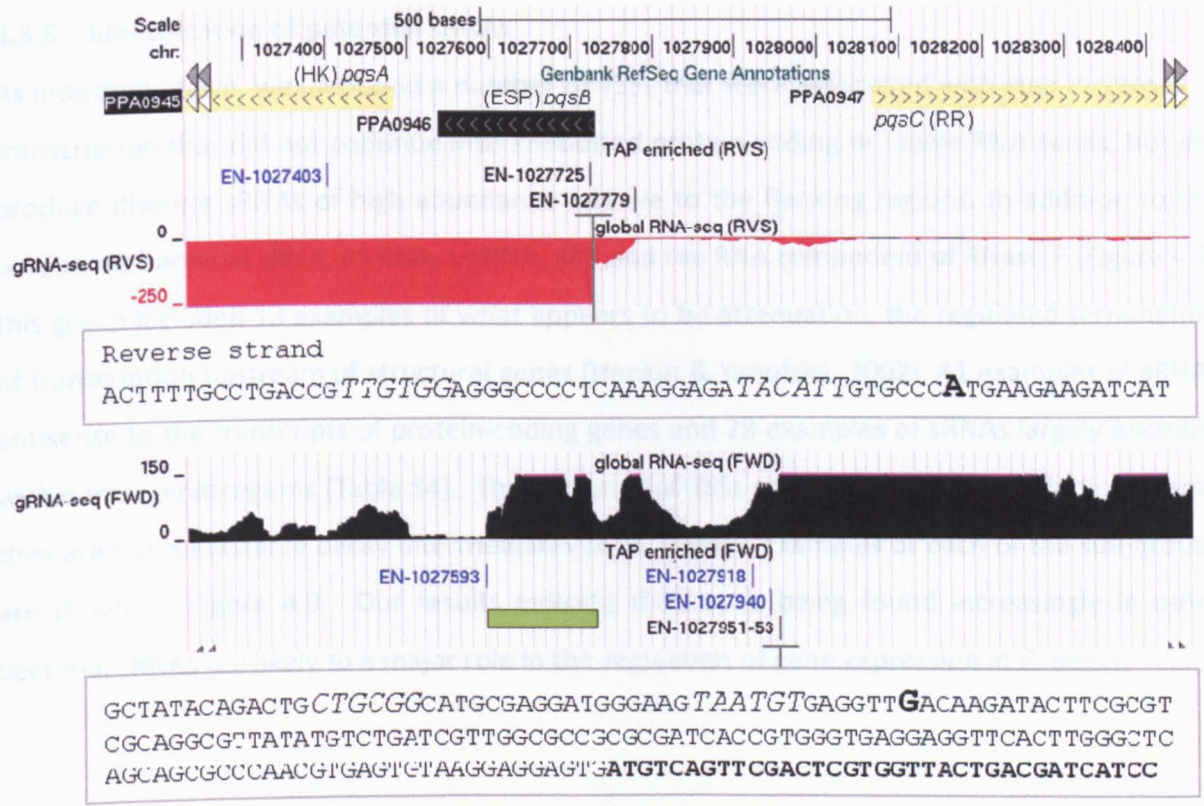


Figure 4.8. RNA-seq analysis of the *pqs* operon. The tracks are as Figure 4.7; remainder of labelling as Figure 4.2. For the insets below the gRNA-seq tracks, large bold fonts indicate transcriptional start sites identified by 5' RACE (Guan, 2011), putative -10 and -35 region are in italic font. The 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*.

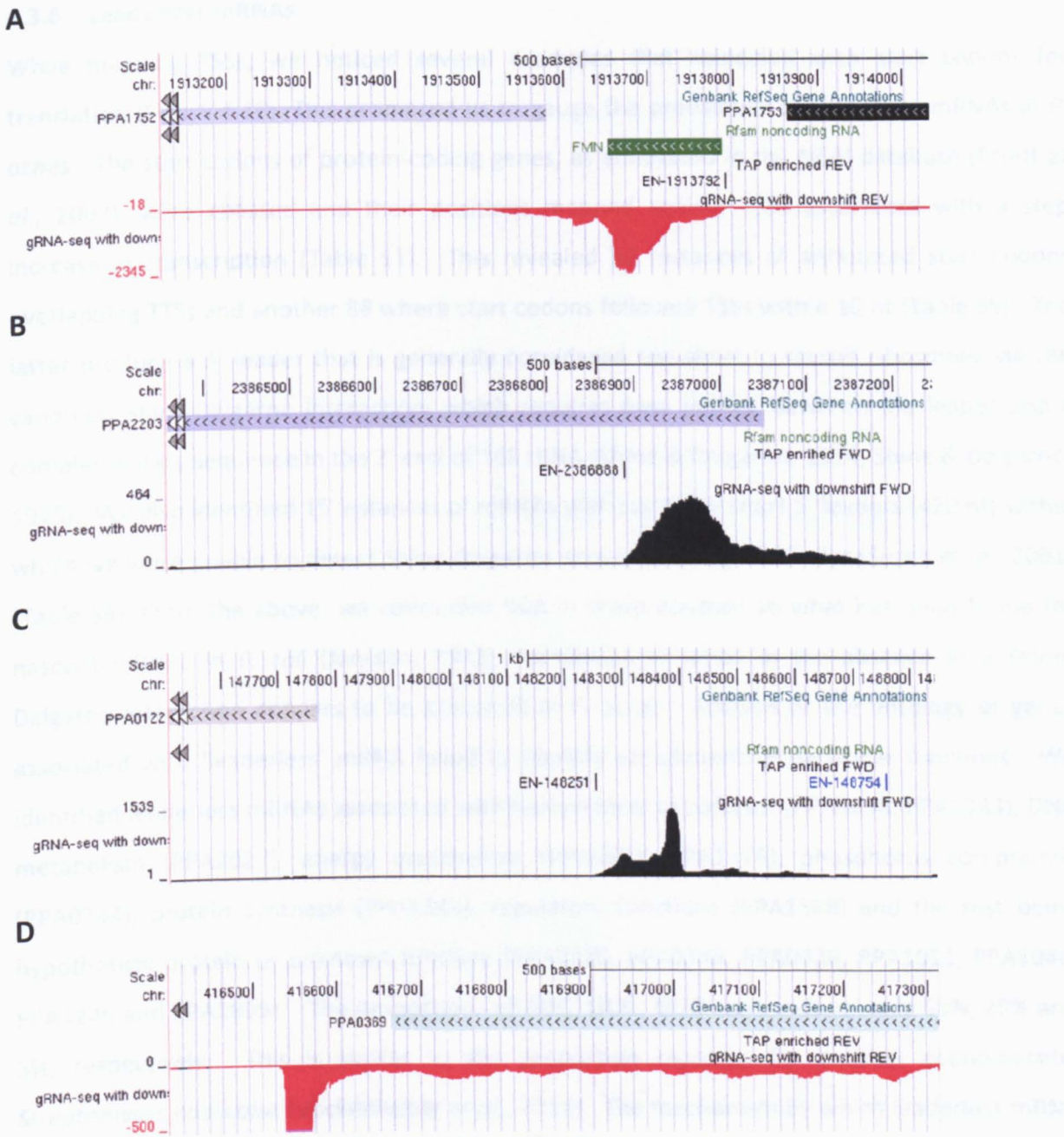


Figure 4.9. Examples of *P. acnes* small RNAs. Panels A, B, C and D correspond to examples of riboswitch, *cis*-encoded antisense RNA, intergenic sRNA of unknown function, and metastable decay intermediates. For each of these panels, the tracks show, from top to bottom, the positions of annotated genes, 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 TTSs 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'-hydoxylated fragments that are produced by toxin mRNases.

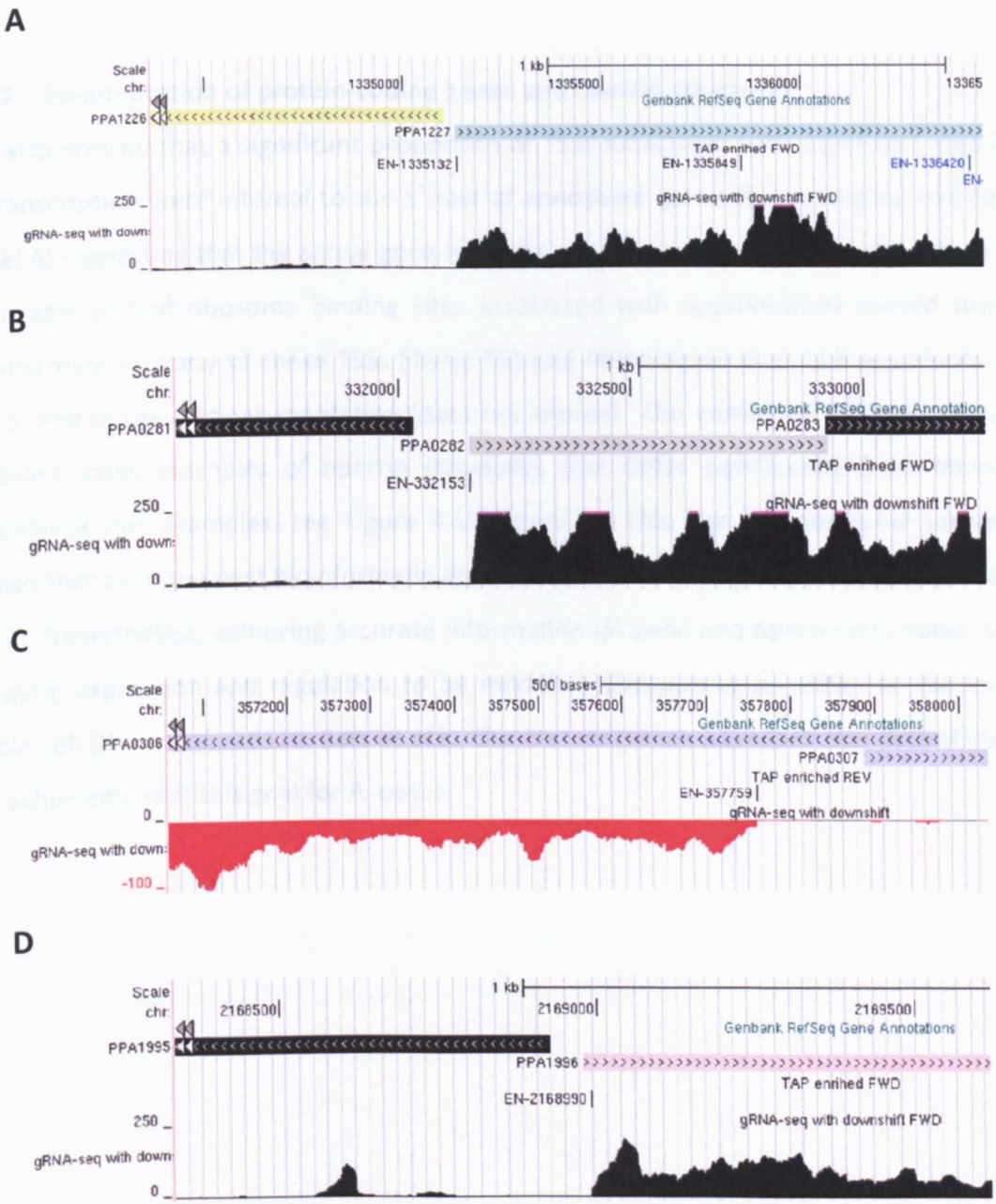


Figure 4.10. Leaderless mRNAs and transcripts of genes requiring reannotation. Panels A and B correspond to examples of leaderless mRNA, while C and D correspond to genes requiring reannotation. For each of these panels, the tracks show, from top to bottom, the positions of annotated genes, the position 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 *cI* 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 *cI* 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 for 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 *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 *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 *P. acnes* genome. This may reflect the addition of 3' tails by the *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 *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, bacteriodes 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, Sörensen *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 3×10^6 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.

Forward strand		Reserve strand
16650	1129092	112423
69338-40	1220499	130158-60
81851	1224843	157289
95813	1248645	387710
133015	1565719	453470-72
145183	1566109	552464
186527	1614866	574372
199776	1654717	578506
218860	1669453	778327
252385	1674593	795754
262666-69	1725349	947658
267845-49	1920360	979898
326577	1933084	1058658
371740	1948896	1163170
382217	1990644	1325267
396972	2066299	1455988
420176	2153965	1583971
476077	2174792	1719382
551864	2269893	1741332
656880-83	2280058	1860031
720186	2294577	1913792
789397	2302239	2038206
892305	2345611	2317705
934712	2386888	
947536	2441486	
951357	2505019	
981381	2538808	
996061	2551263	
1078903		

Table S3. *P. acnes* homologues of genes involved in RNA processing and degradation.

	PPAxxxx	Description
RNase E/G (NTD)	PPA0826	Single-strand-specific endoRNase involved in RNA degradation and the processing of stable RNAs. PPA0826 has extension on N-terminus
RNase III	PPA1452	Double-strand-specific endoRNase involved in processing of rRNA and mRNA. May initiate the cleavage of some mRNAs
RNase P (RNA component)	PPA0652	EndoRNase that processes 5′ end of tRNA. Also processes its own catalytic RNA and cuts some mRNAs
RppH	PPA0342	RNA pyrophosphohydrolase that initiates degradation of some mRNA by hydrolysis of the 5'-triphosphate end
RNase J	PPA1467	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*
RNase Y	PPA1014	Endonuclease involved in the degradation of mRNA in *B. subtilis*, not found in *E. coli*
PNPase	PPA1471	3' to 5' exoRNase and 3'-terminal and oligonucleotide polymerase. Functions in the degradation of various mRNAs and tRNA maturation
oligoRNase	PPA1642	Processive 3'-to-5' exoRNase specific for short oligoribonucleotides. Final enzyme in degraded RNAs to mononucleotides
poly(A) polymerase or tRNA nucleotidyl-transferase	PPA2301	Responsible for oligoadenylation of 3' ends of RNA molecules
RNase PH	PPA1674	3' to 5' exoRNase involved in 3' trimming of tRNAs.
RNase D	PPA1063	3'-5' exoRNase involved in the 3' processing of various stable RNA molecules.
tRNase Z	none	EndoRNase that can generate the mature 3' end of tRNA
nanoRNase	none	Functionally equiavalent to oligoRNase
RNase BN	none	3' to 5' exoRNase involved in 3' trimming of tRNAs as well as various short unstructured RNAs
RNase T	none	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
RNase II and RNase R	none	3′ to 5′ exoRNase that cleaves RNA from the 3' end to produce ribonucleoside 5'-monophosphates
nanoRNase	none	Functionally equivalent to oligoRNase

Table S4. List of annotated and novel sRNAs in *P. acnes*.

Antisense		Riboswitch	Intergenic region	
25506 +	981518 -	16650 +	148251 +	1642083 +
48200 -	1062051 -	33231 +	235787 +	1719380 -
54282 +	1202004 -	95813 +	525729 +	1815486 -
107414 +	1251316 -	199776 +	647993 -	1815487 -
226602 +	1543535 +	218860 +	764909 +	1933084 +
234105 +	1573376 +	326577 +	788348 +	1996134 -
262666 +	1625671 -	476077 +	931846 +	2038081 +
350628 +	1731448 +	578506 -	947536 +	2048808 +
419298 +	1747014 -	1129093 +	1058658 -	2048848 +
429137 -	1892720 +	1163170 -	1248645 +	2050281 +
447936 -	1902900 +	1220499 +	1372793 +	2126592 -
515180 -	2067110 -	1719382 -	1377440 -	2153965 +
523981 +	2154770 -	1913792 -	1566109 +	2321632 -
525730 +	2154776 -	1920361 +	1642082 +	2551421 +
586175 -	2286406 -	1933084 +		
601083 -	2286407 -	1972785 -		
601251 -	2338912 +	2294577 +		
662829 -	2354014 +	2538808 +		
690031 +	2354015 +			
776546 +	2386888 +			
885274 +				

(+/-) 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.

Leaderless mRNA TSS	Strand	Distance from start codon	Gene	Leaderless mRNA TSS	Strand	Distance from start codon	Gene
85492	+	1	PPA0079	1332639	+	2	PPA1225
219005	+	0	PPA0180	1335132	+	0	PPA1227
264693	+	8	PPA0213	1439333	+	0	PPA1323
275563*	+	0	PPA0220	1486763	+	0	PPA1364
301857	+	18	PPA0248	1580884	+	0	PPA1465
310295	+	20	PPA0257	1614866	+	0	PPA1496
312708	+	0	PPA0259	1677186	+	0	PPA1557
332153	+	0	PPA0282	1687656*	+	0	PPA1568
357872	+	15	PPA0307	1789002	+	10	PPA1642
383235	+	0	PPA0333	1851527	+	0	PPA1699
396972	+	9	PPA0345	1875582	+	0	PPA1722
453574	+	0	PPA0408	1924939	+	0	PPA1762
544192	+	0	PPA0494	2070668	+	0	PPA1909
565188	+	0	PPA0513	2103935*	+	1	PPA1943
581080	+	0	PPA0527	2113770	+	0	PPA1953
619022	+	6	PPA0559	2124998	+	20	PPA1962
620292	+	0	PPA0560	2200062*	+	2	PPA2027
632907	+	0	PPA0572	2262758	+	13	PPA2088
636616	+	0	PPA0575	2270276	+	12	PPA2095
649446	+	0	PPA0588	2285027	+	0	PPA2111
651011	+	0	PPA0590	2343229	+	0	PPA2163
701664	+	0	PPA0636	2359454	+	0	PPA2177
707657	+	0	PPA0643	2365576	+	0	PPA2184
712934	+	20	PPA0648	2367840	+	1	PPA2187
719343	+	0	PPA0653	2380556	+	0	PPA2200
720186	+	12	PPA0656	2383635	+	0	PPA2202
744036	+	0	PPA0676	2388012	+	0	PPA2205
795728	+	13	PPA0724	2420039	+	0	PPA2236
802023	+	1	PPA0730	2430005	+	0	PPA2246
843812	+	0	PPA0769	2441486	+	0	PPA2257
977946	+	2	PPA0897	2453150	+	18	PPA2267
1007244	+	0	PPA0924	2484814	+	0	PPA2289
1025484	+	0	PPA0944	2486987	+	0	PPA2292
1042580	+	0	PPA0959				
1094775*	+	15	PPA1011				
1101306	+	3	PPA1017				
1131717	+	0	PPA1040				
1152133	+	2	PPA1060				

Table S5 continued

Leaderless mRNA TSS	Strand	Distance from start codon	Gene	Leaderless mRNA TSS	Strand	Distance from start codon	Gene
157289	-	1	PPA0130	1352429*	-	4	PPA1246
160753	-	1	PPA0133	1364491	-	1	PPA1259
213202	-	1	PPA0176	1378348	-	17	PPA1270
307514	-	19	PPA0252	1468550*	-	1	PPA1344
310324	-	1	PPA0256	1478097	-	1	PPA1354
366974	-	1	PPA0314	1485284	-	16	PPA1360
380907	-	1	PPA0330	1489728	-	1	PPA1366
398318*	-	1	PPA0346	1492227	-	-1	PPA1368
406992	-	17	PPA0357	1497174*	-	7	PPA1376
412518	-	1	PPA0364	1507765	-	8	PPA1387
480388*	-	1	PPA0436	1511081	-	1	PPA1390
527105	-	1	PPA0475	1540932	-	1	PPA1421
578338	-	1	PPA0523	1564809	-	1	PPA1447
583253	-	1	PPA0529	1585728	-	19	PPA1467
621885	-	-1	PPA0561	1633757	-	1	PPA1510
625488	-	1	PPA0564	1652327	-	1	PPA1530
629901	-	1	PPA0569	1654714	-	1	PPA1533
635103	-	1	PPA0573	1669229	-	1	PPA1550
664515	-	18	PPA0599	1762569	-	1	PPA1624
689518	-	1	PPA0623	1766231	-	1	PPA1626
696913	-	1	PPA0630	1787483	-	1	PPA1640
706656	-	1	PPA0641	1833078	-	1	PPA1680
712849	-	1	PPA0647	1849598	-	1	PPA1696
712866	-	18	PPA0647	1851422	-	1	PPA1698
727624*	-	1	PPA0661	1860031	-	1	PPA1705
728897	-	1	PPA0663	1864065	-	1	PPA1710
735523	-	3	PPA0667	1941266	-	1	PPA1776
759849	-	1	PPA0689	1948769	-	1	PPA1785
768211	-	19	PPA0694	1966431	-	1	PPA1801
781355	-	1	PPA0708	2057156	-	1	PPA1896
817762*	-	1	PPA0744	2060023*	-	10	PPA1899
924594	-	1	PPA0845	2062981	-	1	PPA1903
1069245	-	1	PPA0986	2062988	-	8	PPA1903
1141827*	-	1	PPA1046	2112549	-	0	PPA1951
1146258	-	1	PPA1052	2131829	-	11	PPA1968
1160415	-	13	PPA1064	2134455	-	1	PPA1970
1196203	-	1	PPA1104	2145840	-	1	PPA1977
1198468	-	1	PPA1105	2160342	-	1	PPA1988

Table S5 continued

Leaderless mRNA TSS	Strand	Distance from start codon	Gene
2174714	-	1	PPA2002
2181571	-	1	PPA2008
2187838	-	1	PPA2015
2289917	-	1	PPA2115
2305442	-	1	PPA2128
2311086	-	1	PPA2133
2410592	-	1	PPA2225
2419928	-	1	PPA2235
2427001	-	13	PPA2243
2436830	-	0	PPA2251
2493898	-	1	PPA2297

*TSSs where no Shine-Dalgarno sequence was found using RBSfinder.

Table S6. List of genes requiring reannotation.

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

Table S7. Composition of Holland defined medium

Solution I [a,b]

Components	Per 100ml (mg)
L-Alanine	10
L-Arginine	20
L-Asparagine	10
L-Aspartic acid	20
L-Glutamic acid	50
Glycine	10
L-Histidine	10
L-Isoleucine	10
L-Hydroxyproline	10
L-Cysteine	10
L-Lysine	10
L-Leucine	10
L-Phenylalanine	10
L-Methionine	10
L-Proline	10
L-Serine	20
L-Threonine	20
L-Tryptophan	10
L-Tyrosin	10
L-Valine	10
Glucose	500
PIPES	1000
K_2HPO_4	350
KH_2PO_4	350
$MgSO_4.7H_2O$	20
NaCl	200
Sodium citrate	50
$(NH4)_2SO_4$	150

Mineral salt components[c†]	Cocentration (g/100ml)	Vitamin components[c‡]	Concentration (g/100ml)
$FeSO_4.7H_2O$	1.0	Pyridoxine HCl	1.25
$MnSO_4.4H_2O$	1.0	Nicotinic acid	0.25
$CaCl_2.2H_2O$	1.0	Pantothenic acid	0.125
$ZnCl_2$	0.25	Biotin	0.015
$CoCl_2.6H_2O$	0.25		
$CuSO_4.5H_2O$	0.05		

† - Mineral salts were dissolved in 1.0M HCl.

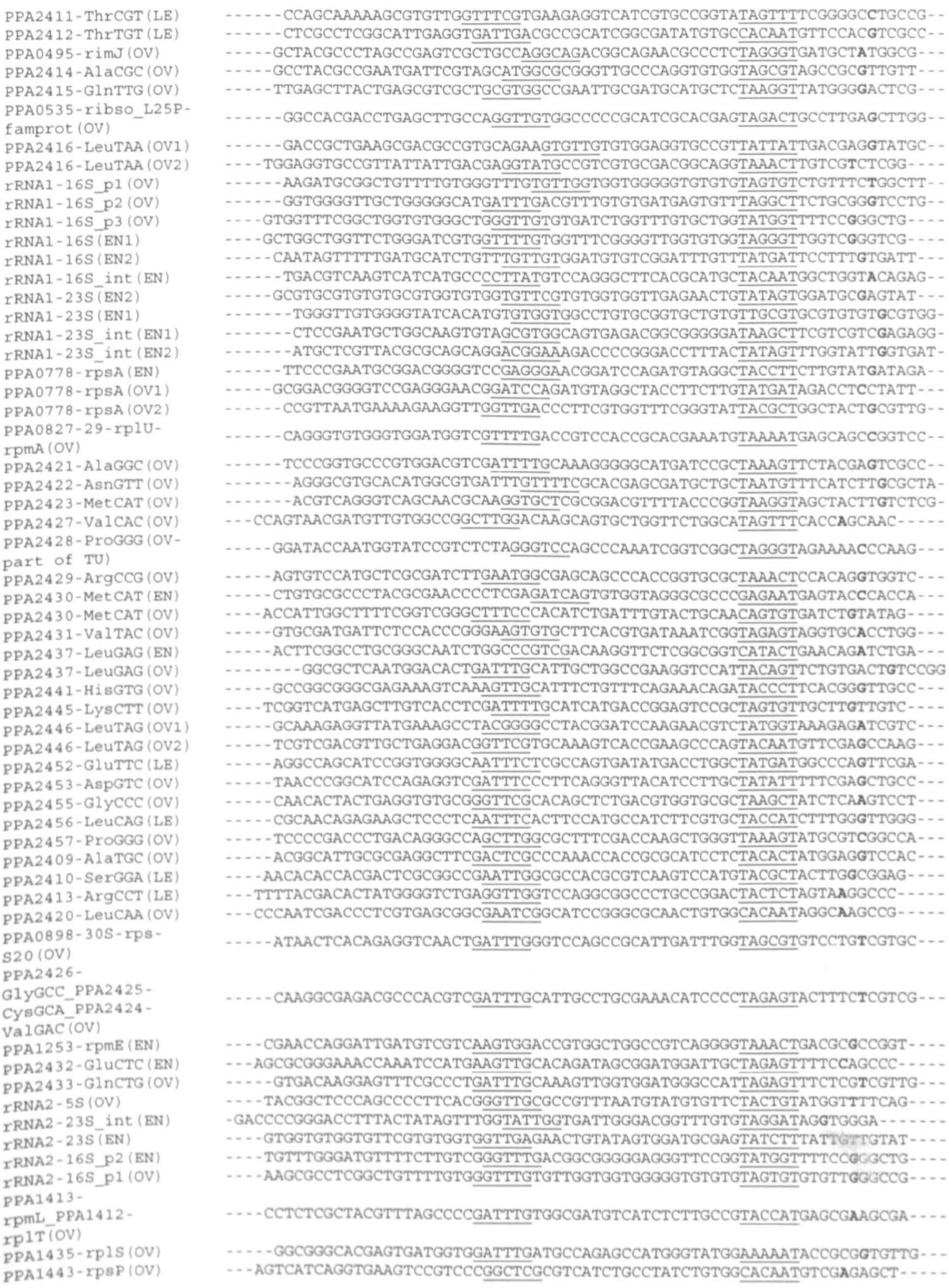
‡ - Vitamins were dissolved in dH_2O

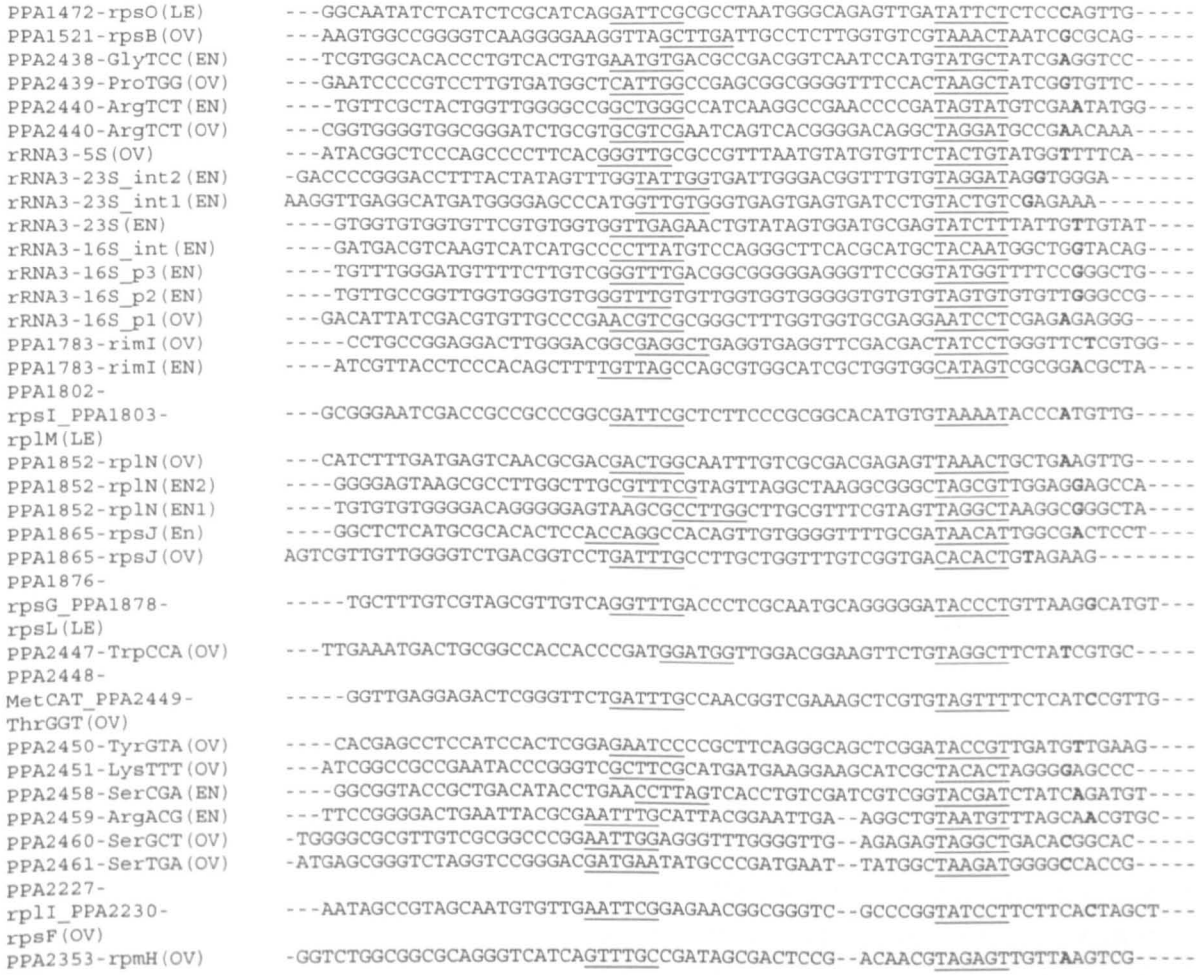
a - Tween-80 added at 0.1% [v/v] of Solution I

b - final pH of HSM is 5.6

c - Mineral salt and vitamin solutions added at 0.2% [v/v] of Solution I

A


PPA2411-ThrCGT(LE)
PPA2412-ThrTGT(LE)
PPA0495-rimJ(OV)
PPA2414-AlaCGC(OV)
PPA2415-GlnTTG(OV)
PPA0535-ribso_L25P-famprot(OV)
PPA2416-LeuTAA(OV1)
PPA2416-LeuTAA(OV2)
rRNA1-16S_p1(OV)
rRNA1-16S_p2(OV)
rRNA1-16S_p3(OV)
rRNA1-16S(EN1)
rRNA1-16S(EN2)
rRNA1-16S_int(EN)
rRNA1-23S(EN2)
rRNA1-23S(EN1)
rRNA1-23S_int(EN1)
rRNA1-23S_int(EN2)
PPA0778-rpsA(EN)
PPA0778-rpsA(OV1)
PPA0778-rpsA(OV2)
PPA0827-29-rplU-rpmA(OV)
PPA2421-AlaGGC(OV)
PPA2422-AsnGTT(OV)
PPA2423-MetCAT(OV)
PPA2427-ValCAC(OV)
PPA2428-ProGGG(OV-part of TU)
PPA2429-ArgCCG(OV)
PPA2430-MetCAT(EN)
PPA2430-MetCAT(OV)
PPA2431-ValTAC(OV)
PPA2437-LeuGAG(EN)
PPA2437-LeuGAG(OV)
PPA2441-HisGTG(OV)
PPA2445-LysCTT(OV)
PPA2446-LeuTAG(OV1)
PPA2446-LeuTAG(OV2)
PPA2452-GluTTC(LE)
PPA2453-AspGTC(OV)
PPA2455-GlyCCC(OV)
PPA2456-LeuCAG(LE)
PPA2457-ProGGG(OV)
PPA2409-AlaTGC(OV)
PPA2410-SerGGA(LE)
PPA2413-ArgCCT(LE)
PPA2420-LeuCAA(OV)
PPA0898-30S-rps-S20(OV)
PPA2426-GlyGCC_PPA2425-CysGCA_PPA2424-ValGAC(OV)
PPA1253-rpmE(EN)
PPA2432-GluCTC(EN)
PPA2433-GlnCTG(OV)
rRNA2-5S(OV)
rRNA2-23S_int(EN)
rRNA2-23S(EN)
rRNA2-16S_p2(EN)
rRNA2-16S_p1(OV)
PPA1413-rpmL_PPA1412-rplT(OV)
PPA1435-rplS(OV)
PPA1443-rpsP(OV)


PPA1472-rpsO(LE)
PPA1521-rpsB(OV)
PPA2438-GlyTCC(EN)
PPA2439-ProTGG(OV)
PPA2440-ArgTCT(EN)
PPA2440-ArgTCT(OV)
rRNA3-5S(OV)
rRNA3-23S_int2(EN)
rRNA3-23S_int1(EN)
rRNA3-23S(EN)
rRNA3-16S_int(EN)
rRNA3-16S_p3(EN)
rRNA3-16S_p2(EN)
rRNA3-16S_p1(OV)
PPA1783-rimI(OV)
PPA1783-rimI(EN)
PPA1802-
rpsI_PPA1803-
rplM(LE)
PPA1852-rplN(OV)
PPA1852-rplN(EN2)
PPA1852-rplN(EN1)
PPA1865-rpsJ(En)
PPA1865-rpsJ(OV)
PPA1876-
rpsG_PPA1878-
rpsL(LE)
PPA2447-TrpCCA(OV)
PPA2448-
MetCAT_PPA2449-
ThrGGT(OV)
PPA2450-TyrGTA(OV)
PPA2451-LysTTT(OV)
PPA2458-SerCGA(EN)
PPA2459-ArgACG(EN)
PPA2460-SerGCT(OV)
PPA2461-SerTGA(OV)
PPA2227-
rplI_PPA2230-
rpsF(OV)
PPA2353-rpmH(OV)

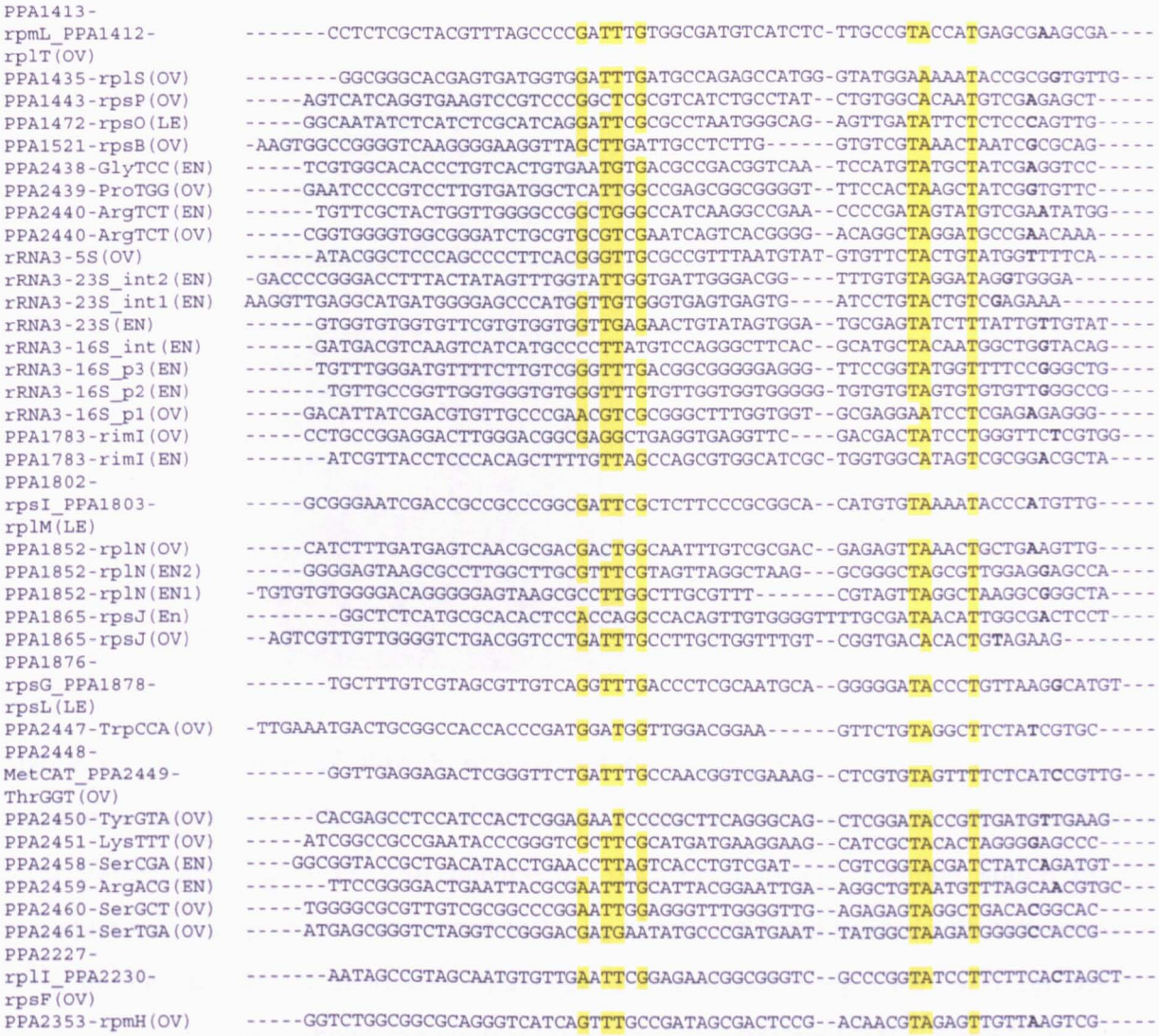


Figure S1. Sequence alignment of promoters associated with the translational machinery. Panel A shows ungapped sequences (+5 to -60) aligned to the '-10 box' (consensus sequence of TAnnnT), which was identified using MEME (Bailey & Gribskov, 1998) and an initial search window of -1 to -15. Panel B as 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.

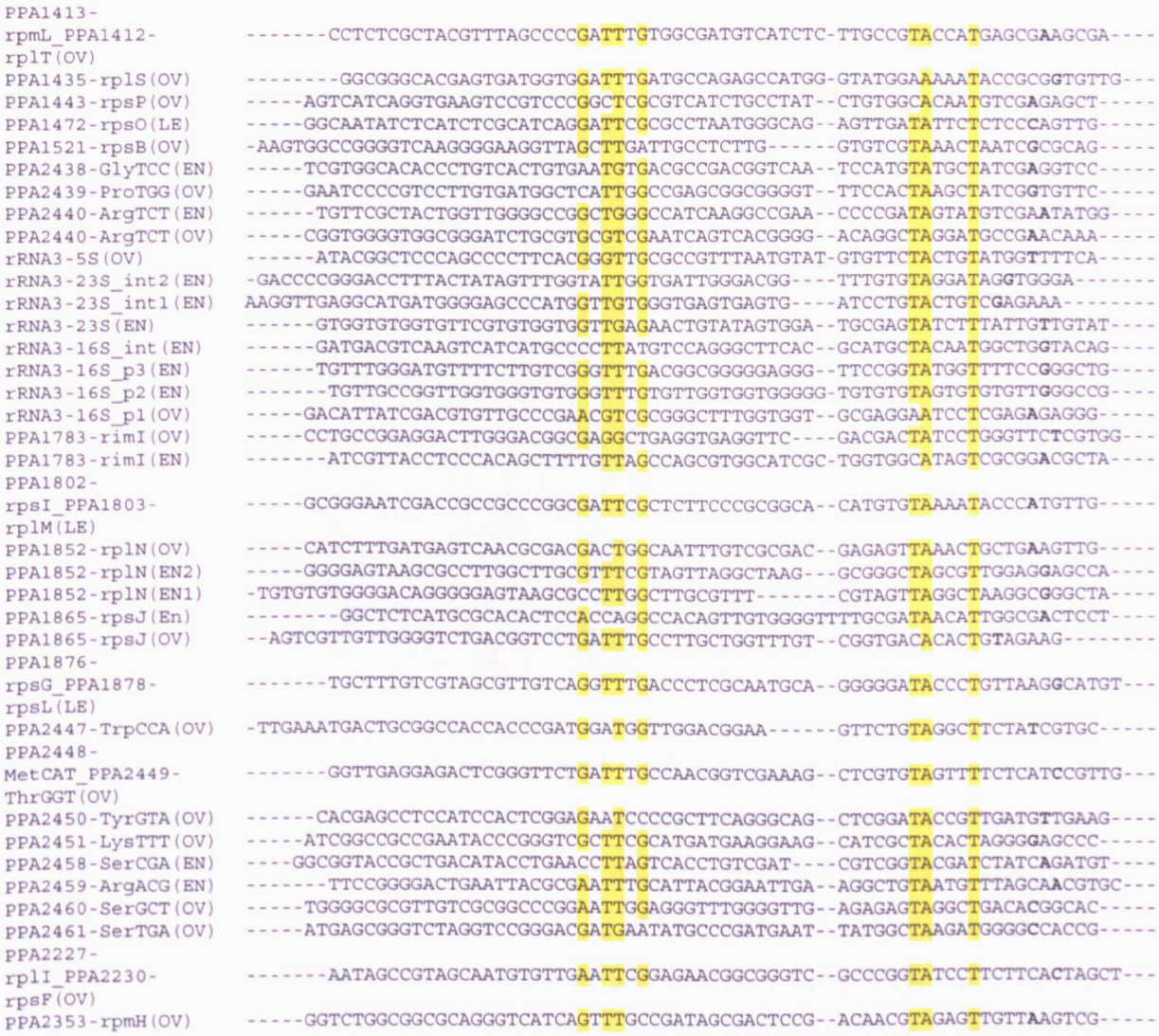


Figure S1. Sequence alignment of promoters associated with the translational machinery. Panel A shows ungapped sequences (+5 to -60) aligned to the '-10 box' (consensus sequence of TAnnnT), which was identified using MEME (Bailey & Gribskov, 1998) and an initial search window of -1 to -15. Panel B as 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.

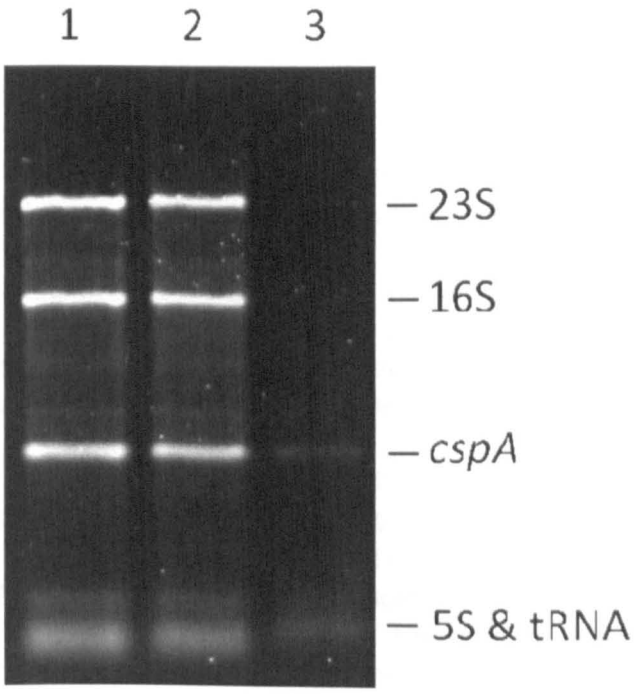


Figure S2. Degradation of 5'-triphosphorylated RNA using TEX. Total *P. acnes* RNA was isolated as described in Materials and Methods. *E. coli cspA* transcript was synthesised by *in vitro* transcription using T7 RNA polymerase (Invitrogen) using condition stated by the manufacturer. 0.5 µg of *cspA* 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] agroase 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

Adebamowo, C. A., D. Spiegelman, C. S. Berkey, F. W. Danby, H. H. Rockett, G. A. Colditz, W. C. Willett & M. D. Holmes, (2008) Milk consumption and acne in teenaged boys. *J Am Acad Dermatol* **58**: 787-793.

Albrecht, M., C. M. Sharma, R. Reinhardt, J. Vogel & T. Rudel, (2010) Deep sequencing-based discovery of the *Chlamydia trachomatis* transcriptome. *Nucleic Acids Res.* **38**: 868-877.

Altendorf, K., P. Voelkner & W. Puppe, (1994) The sensor kinase KdpD and the response regulator KdpE control expression of the *kdpFABC* operon in *Escherichia coli*. *Res Microbiol* **145**: 374-381.

Andrade, J. M., V. Pobre, I. J. Silva, S. Domingues & C. M. Arraiano, (2009) The role of 3 '-5 ' exoribonucleases in RNA degradation. In: Molecular Biology of Rna Processing and Decay in Prokaryotes. C. Condon (ed). pp. 187-229.

Arraiano, C. M., J. M. Andrade, S. Domingues, I. B. Guinote, M. Malecki, R. G. Matos, R. N. Moreira, V. Pobre, F. P. Reis, M. Saramago, I. J. Silva & S. C. Viegas, (2010) The critical role of RNA processing and degradation in the control of gene expression. *FEMS Microbiol Rev* **34**: 883-923.

Baba, T., T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. A. Datsenko, M. Tomita, B. L. Wanner & H. Mori, (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**: 2006 0008.

Bailey, T. L., M. Boden, F. A. Buske, M. Frith, C. E. Grant, L. Clementi, J. Y. Ren, W. W. Li & W. S. Noble, (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**: W202-W208.

Bailey, T. L. & M. Gribskov, (1998) Combining evidence using p-values: application to sequence homology searches. *Bioinformatics* **14**: 48-54.

Ballal, A. & S. K. Apte, (2005) Differential expression of the two *kdp* operons in the nitrogen-fixing cyanobacterium *Anabaena sp.* strain L-31. *Appl Environ Microbiol* **71**: 5297-5303.

Ballal, A., B. Basu & S. K. Apte, (2007) The Kdp-ATPase system and its regulation. *J Biosci* **32**: 559-568.

Balmer, J. E. & R. Blomhoff, (2002) Gene expression regulation by retinoic acid. *J Lipid Res* **43**: 1773-1808.

Banin, E., M. L. Vasil & E. P. Greenberg, (2005) Iron and *Pseudomonas aeruginosa* biofilm formation. *Proc Natl Acad Sci USA* **102**: 11076-11081.

Barrick, J. E. & R. R. Breaker, (2007) The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biol.* **8**.

Baumeister, R., P. Flache, O. Melefors, A. Vongabain & W. Hillen, (1991) Lack of a 5' noncoding region in Tn1721-encoded *tetR* mRNA is associated with a low efficiency of translation and a short half-life in *Escherichia coli*. *Nucleic Acids Res.* **19**: 4595-4600.

Bayston, R., W. Ashraf, R. Barker-Davies, E. Tucker, R. Clement, J. Clayton, B. J. C. Freeman & B. Nuradeen, (2007) Biofilm formation by *Propionibacterium acnes* on biomaterials *in vitro* and *in vivo*: Impact on diagnosis and treatment. *J Biomed Mater Res* **81A**: 705-709.

Beaume, M., D. Hernandez, L. Farinelli, C. Deluen, P. Linder, C. Gaspin, P. Romby, J. Schrenzel & P. Francois, (2010) Cartography of methicillin-resistant *S. aureus* transcripts: detection, orientation and temporal expression during growth phase and stress conditions. *Plos One* **5**.

Bechhofer, D. H., (2009) Messenger RNA Decay and Maturation in *Bacillus subtilis*. *Molecular Biology of RNA Processing and Decay in Prokaryotes* **85**: 231-273.

Belasco, J. G., (2010) All things must pass: contrasts and commonalities in eukaryotic and bacterial mRNA decay. *Nat. Rev. Mol. Cell. Bio.* **11**: 467-478.

Berk, A. J. & P. A. Sharp, (1977) Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. *Cell* **12**: 721-732.

Bier, F. F. & F. Kleinjung, (2001) Feature-size limitations of microarray technology - a critical review. *Fresenius' journal of analytical chemistry* **371**: 151-156.

Bojar, R. A. & K. T. Holland, (2004) Acne and *Propionibacterium acnes*. *Clin Dermatol* **22**: 375-379.

Borovok, I., B. Gorovitz, M. Yanku, R. Schreiber, B. Gust, K. Chater, Y. Aharonowitz & G. Cohen, (2004) Alternative oxygen-dependent and oxygen-independent ribonucleotide reductases in *Streptomyces*: cross-regulation and physiological role in response to oxygen limitation. *Mol. Microbiol.* **54**: 1022-1035.

Breitling, R., P. Armengaud, A. Amtmann & P. Herzyk, (2004) Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS Lett* **573**: 83-92.

Breitling, R. & P. Herzyk, (2005) Rank-based methods as a non-parametric alternative of the T-statistic for the analysis of biological microarray data. *J Bioinform Comput Biol* **3**: 1171-1189.

Breter, H. J. & R. E. Rhoads, (1979) Analysis of $NaIO_4$-oxidized/$NaBH_4$-reduced mRNA cap analogs by high-performance liquid anion-exchange chromatography and tobacco acid pyrophosphatase (EC 3.6.1.9). *H-S Physiol. Chem.* **360**: 240-240.

Brock, J. E., S. Pourshahian, J. Giliberti, P. A. Limbach & G. R. Janssen, (2008) Ribosomes bind leaderless mRNA in *Escherichia coli* through recognition of their 5 '-terminal AUG. *RNA* **14**: 2159-2169.

Brogden, K. A. & J. M. Guthmiller, (2002) *Polymicrobial diseases*. ASM Press.

Brüggemann, H., (2005) Insights in the pathogenic potential of *Propionibacterium acnes* from its complete genome. In: Semin Cutan Med Surg. Elsevier, pp. 67-72.

Bruggemann, H., A. Henne, F. Hoster, H. Liesegang, A. Wiezer, A. Strittmatter, S. Hujer, P. Durre & G. Gottschalk, (2004) The complete genome sequence of *Propionibacterium acnes*, a commensal of human skin. *Science* **305**: 671-673.

Brzuszkiewicz, E., J. Weiner, A. Wollherr, A. Thurmer, J. Hupeden, H. B. Lomholt, M. Kilian, G. Gottschalk, R. Daniel, H. J. Mollenkopf, T. F. Meyer & H. Bruggemann, (2011) Comparative genomics and transcriptomics of *Propionibacterium acnes*. *PLoS One* **6**: e21581.

Burkhart, C. N. & C. G. Burkhart, (2003) Microbiology's principle of biofilms as a major factor in the pathogenesis of acne vulgaris. *Int J Dermatol* **42**: 925-927.

Butler-Wu, S. M., E. M. Burns, P. S. Pottinger, A. S. Magaret, J. L. Rakeman, F. A. Matsen, 3rd & B. T. Cookson, (2011) Optimization of periprosthetic culture for diagnosis of *Propionibacterium acnes* prosthetic joint infection. *J Clin Microbiol* **49**: 2490-2495.

Calandra, G. B. & R. M. Cole, (1980) Lysis and protoplast formation of group B streptococci by mutanolysin. *Infect Immun* **28**: 1033-1037.

Camisa, C., B. Eisenstat, A. Ragaz & G. Weissmann, (1982) The effects of retinoids on neutrophil functions *in vitro*. *J Am Acad Dermatol* **6**: 620-629.

Cappel, M., D. Mauger & D. Thiboutot, (2005) Correlation between serum levels of insulin-like growth factor 1, dehydroepiandrosterone sulfate, and dihydrotestosterone and acne lesion counts in adult women. *Arch Dermatol* **141**: 333-338.

Capra, E. J. & M. T. Laub, (2012) Evolution of two-component signal transduction systems. *Annu Rev Microbiol* **66**: 325-347.

Carpousis, A. J., B. F. Luisi & K. J. McDowall, (2009a) Endonucleolytic initiation of mRNA decay in *Escherichia coli*. *Progress in Molecular Biology and Translational Science* **85**: 91-135.

Carpousis, A. J., B. F. Luisi & K. J. McDowall, (2009b) Endonucleolytic Initiation of mRNA Decay in *Escherichia coli*. *Molecular Biology of RNA Processing and Decay in Prokaryotes* **85**: 91-135.

Celesnik, H., A. Deana & J. G. Belasco, (2007) Initiation of RNA decay in *Escherichia coli* by 5 ' pyrophosphate removal. *Mol. Cell* **27**: 79-90.

Challis, G. L. & J. H. Naismith, (2004) Structural aspects of non-ribosomal peptide biosynthesis. *Curr Opin Struct Biol* **14**: 748-756.

Chaudhuri, R. R., S. E. Peters, S. J. Pleasance, H. Northen, C. Willers, G. K. Paterson, D. B. Cone, A. G. Allen, P. J. Owen, G. Shalom, D. J. Stekel, I. G. Charles & D. J. Maskell, (2009) Comprehensive identification of *Salmonella enterica* serovar Typhimurium genes required for infection of BALB/c mice. *PLoS pathogens* **5**: e1000529.

Chauhan, A. K. & D. Apirion, (1989) The gene for a small stable RNA (10Sa RNA) of *Escherichia coli*. *Mol. Microbiol.* **3**: 1481-1485.

Cheong, D.-E., H.-I. Lee & J.-S. So, (2008) Optimization of electrotransformation conditions for *Propionibacterium acnes*. *J Microbiol Methods* **72**: 38-41.

Cho, B. K., K. Zengler, Y. Qiu, Y. S. Park, E. M. Knight, C. L. Barrett, Y. Gao & B. O. Palsson, (2009) The transcription unit architecture of the *Escherichia coli* genome. *Nat. Biotech.* **27**: 1043-U1115.

Cholo, M. C., R. Anderson & E. J. van Rensburg, (2009) Potassium uptake systems of *Mycobacterium tuberculosis*: genomic and protein organisation and potential roles in microbial pathogenesis and chemotherapy. *South Afr J Epidemiol Infect* **23**.

Chuang, L. Y., H. W. Chang, J. H. Tsai & C. H. Yang, (2012) Features for computational operon prediction in prokaryotes. *Briefings in Functional Genomics* **11**: 291-299.

Claesson, M. J., I. B. Jeffery, S. Conde, S. E. Power, E. M. O'Connor, S. Cusack, H. M. Harris, M. Coakley, B. Lakshminarayanan, O. O'Sullivan, G. F. Fitzgerald, J. Deane, M. O'Connor, N. Harnedy, K. O'Connor, D. O'Mahony, D. van Sinderen, M. Wallace, L. Brennan. C. Stanton, J. R. Marchesi, A. P. Fitzgerald, F. Shanahan, C. Hill, R. P. Ross & P. W. O'Toole, (2012) Gut microbiota composition correlates with diet and health in the elderly. *Nature* **488**: 178-184.

Coates, P., S. Vyakrnam, E. Eady, C. Jones, J. Cove & W. Cunliffe, (2002) Prevalence of antibiotic-resistant propionibacteria on the skin of acne patients: 10-year surveillance data and snapshot distribution study. *Br J Dermatol* **146**: 840-848.

Coenye, T., G. Brackman, P. Rigole, E. De Witte, K. Honraet, B. Rossel & H. J. Nelis, (2012) Eradication of *Propionibacterium acnes* biofilms by plant extracts and putative identification of icariin, resveratrol and salidroside as active compounds. *Phytomedicine* **19**: 409-412.

Coenye, T., E. Peeters & H. J. Nelis, (2007) Biofilm formation by *Propionibacterium acnes* is associated with increased resistance to antimicrobial agents and increased production of putative virulence factors. *Res Microbiol* **158**: 386-392.

Costerton, J. W., Z. Lewandowski, D. E. Caldwell, D. R. Korber & H. M. Lappin-Scott, (1995) Microbial biofilms. *Annu Rev Microbiol* **49**: 711-745.

Cove, J., W. Cunliffe & K. Holland, (2006) Acne vulgaris: is the bacterial population size significant? *Br J Dermatol* **102**: 277-280.

Cove, J. H., K. T. Holland & W. J. Cunliffe, (1983) Effects of oxygen concentration on biomass production, maximum specific growth rate and extracellular enzyme production by three species of cutaneous propionibacteria grown in continuous culture. *J Gen Microbiol* **129**: 3327-3334.

Cox, J. & M. Mann, (2011) Quantitative, high-resolution proteomics for data-driven systems biology. *Annu Rev Biochem* **80**: 273-299.

Crooks, G. E., G. Hon, J.-M. Chandonia & S. E. Brenner, (2004) WebLogo: a sequence logo generator. *Genome Res* **14**: 1188-1190.

Csonka, L. N. & A. D. Hanson, (1991) Prokaryotic osmoregulation: genetics and physiology. *Annu Rev Microbiol* **45**: 569-606.

Cunliffe, W. J., (1986) Acne and unemployment. *Br J Dermatol* **115**: 386.

Cunliffe, W. J., D. B. Holland & A. Jeremy, (2004) Comedone formation: etiology, clinical presentation, and treatment. *Clin Dermatol* **22**: 367-374.

Danby, F. W., (2010) Nutrition and acne. *Clin Dermatol* **28**: 598-604.

de Beer, D., P. Stoodley, F. Roe & Z. Lewandowski, (1994) Effects of biofilm structures on oxygen distribution and mass transport. *Biotechnol Bioeng* **43**: 1131-1138.

de Boer, H. A., S. F. Gilbert & M. Nomura, (1979) DNA sequences of promoter regions for rRNA operons *rrnE* and *rrnA* in *Escherichia coli*. *Cell* **17**: 201-209.

De Keersmaecker, S. C., I. M. Thijs, J. Vanderleyden & K. Marchal, (2006) Integration of omics data: how well does it work for bacteria? *Mol Microbiol* **62**: 1239-1250.

Deana, A., H. Celesnik & J. G. Belasco, (2008) The bacterial enzyme RppH triggers messenger RNA degradation by 5' pyrophosphate removal. *Nature* **451**: 355-U314.

Deutscher, M. P., (2009) Maturation and degradation of ribosomal RNA in bacteria. *Molecular Biology of RNA Processing and Decay in Prokaryotes* **85**: 369-391.

Dittmar, K. A., E. M. Mobley, A. J. Radek & T. Pan, (2004) Exploring the regulation of tRNA distribution on the genomic scale. *J. Mol. Biol.* **337**: 31-47.

Dodson, C. C., E. V. Craig, F. A. Cordasco, D. M. Dines, J. S. Dines, E. Dicarlo, B. D. Brause & R. F. Warren, (2010) *Propionibacterium acnes* infection after shoulder arthroplasty: a diagnostic challenge. *J Shoulder Elbow Surg* **19**: 303-307.

Dornenburg, J. E., A. M. DeVita, M. J. Palumbo & J. T. Wade, (2010) Widespread antisense transcription in *Escherichia coli*. *Mbio* **1**.

Dotsch, A., D. Eckweiler, M. Schniederjans, A. Zimmermann, V. Jensen, M. Scharfe, R. Geffers & S. Haussler, (2012) The *Pseudomonas aeruginosa* transcriptome in planktonic cultures and static biofilms using RNA sequencing. *PLoS One* **7**: e31092.

Downing, D. T., M. E. Stewart, P. W. Wertz & J. S. Strauss, (1986) Essential fatty acids and acne. *J Am Acad Dermatol* **14**: 221-225.

Eady, E., J. Cove, K. Holland & W. Cunliffe, (2006) Erythromycin resistant propionibacteria in antibiotic treated acne patients: association with therapeutic failure. *Br J Dermatol* **121**: 51-57.

Eady, E. A., M. Gloor & J. J. Leyden, (2003) *Propionibacterium acnes* resistance: a worldwide problem. *Dermatology* **206**: 54-56.

Eleaume, H. & S. Jabbouri, (2004) Comparison of two standardisation methods in real-time quantitative RT-PCR to follow *Staphylococcus aureus* genes expression during *in vitro* growth. *J Microbiol Methods* **59**: 363-370.

Epstein, W., (2003) The roles and regulation of potassium in bacteria. *Prog Nucleic Acid Re* **75**: 293-320.

Even, S., O. Pellegrini, L. Zig, V. Labas, J. Vinh, D. Brechemmier-Baey & H. Putzer, (2005) Ribonucleases J1 and J2: two novel endoribonucleases in *B. subtilis* with functional homology to *E. coli* RNase E. *Nucleic Acids Res.* **33**: 2141-2152.

Fabret, C., V. A. Feher & J. A. Hoch, (1999) Two-component signal transduction in *Bacillus subtilis*: how one organism sees its world. *J Bacteriol* **181**: 1975-1983.

Fan, X., Y. Z. Xing, L. H. Liu, C. Liu, D. D. Wang, R. Y. Yang & M. Lapidoth, (2012) Effects of 420-nm intense pulsed light in an acne animal model. *J Eur Acad Dermatol Venereol.*

Farrar, M. D., K. M. Howson, R. A. Bojar, D. West, J. C. Towler, J. Parry, K. Pelton & K. T. Holland, (2007) Genome sequence and analysis of a *Propionibacterium acnes* bacteriophage. *J Bacteriol* **189**: 4161-4167.

Farrar, M. D. & E. Ingham, (2004) Acne: inflammation. *Clin Dermatol* **22**: 380-384.

Filiatrault, M. J., P. V. Stodghill, P. A. Bronstein, S. Moll, M. Lindeberg, G. Grills, P. Schweitzer, W. Wang, G. P. Schroth, S. J. Luo, I. Khrebtukova, Y. Yang, T. Thannhauser, B. G. Butcher, S. Cartinhour & D. J. Schneider, (2010) Transcriptome analysis of *Pseudomonas syringae* identifies new genes, noncoding RNAs, and antisense activity. *J. Bacteriol.* **192**: 2359-2372.

Forth, T., (2012) Reconstruction, visualisation and analysis of an experimentally parameterised metabolic model of the human acute malaria parasite *Plasmodium falciparum*. In.: University of Leeds, pp.

Franklund, C. V. & R. J. Kadner, (1997) Multiple transcribed elements control expression of the *Escherichia coli btuB* gene. *J. Bacteriol.* **179**: 4039-4042.

Freinkel, R. K. & Y. Shen, (1969) The origin of free fatty acids in sebum. II. Assay of the lipases of the cutaneous bacteria and effects of pH. *J Invest Dermatol* **53**: 422-427.

Frymier, J. S., T. D. Reed, S. A. Fletcher & L. N. Csonka, (1997) Characterization of transcriptional regulation of the *kdp* operon of *Salmonella typhimurium*. *J Bacteriol* **179**: 3061-3063.

Furste, J. P., W. Pansegrau, G. Ziegelin, M. Kroger & E. Lanka, (1989) Conjugative transfer of promiscuous IncP plasmids: interaction of plasmid-encoded products with the transfer origin. *Proc Natl Acad Sci USA* **86**: 1771-1775.

Gadd, G. M., (1990) Heavy metal accumulation by bacteria and other microorganisms. *Cell Mol Life Sci* **46**: 834-840.

Gassel, M., T. Mollenkamp, W. Puppe & K. Altendorf, (1999) The KdpF subunit is part of the K^+-translocating Kdp complex of *Escherichia coli* and is responsible for stabilization of the complex *in vitro*. *J Biol Chem* **274**: 37901-37907.

Gatewood, M. L., P. Bralley & G. H. Jones, (2011) RNase III-dependent expression of the *rpsO-pnp* operon of *Streptomyces coelicolor*. *J. Bacteriol.* **193**: 4371-4379.

Georg, J., B. Voss, I. Scholz, J. Mitschke, A. Wilde & W. R. Hess, (2009) Evidence for a major role of antisense RNAs in cyanobacterial gene regulation. *Molecular Systems Biology* **5**.

Ghora, B. K. & D. Apirion, (1978) Structural analysis and *in vitro* processing to $_{p}$5 rRNA of a 9S RNA molecule isolated from an *rne* mutant of *E. coli*. *Cell* **15**: 1055-1066.

Gilbert, S. F., H. A. de Boer & M. Nomura, (1979) Identification of initiation sites for the *in vitro* transcription of rRNA operons *rrnE* and *rrnA* in *Escherichia coli*. *Cell* **17**: 211-224.

Glass, C. K. & M. G. Rosenfeld, (2000) The coregulator exchange in transcriptional functions of nuclear receptors. *Genes Dev* **14**: 121-141.

Glynn, B., K. Lacey, P. Palta, L. Kaplinski, M. Remm, T. Barry, T. Smith & M. Maher, (2007) Demonstration of the application of the tmRNA transcript of the bacterial *ssrA* gene as a molecular diagnostic target using a combination of NASBA and BiaCore technologies. *International Journal of Antimicrobial Agents* **29**: S392-S392.

Goecks, J., A. Nekrutenko, J. Taylor & G. Team, (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* **11**.

Gold, L., (1988) Posttranscriptional regulatory mechanisms in *Escherichia coli*. *Annu Rev Biochem* **57**: 199-233.

Goldsmith, Z. G. & N. Dhanasekaran, (2004) The microrevolution: applications and impacts of microarray technology on molecular biology and medicine (review). *Int J Mol Med* **13**: 483.

Gollnick, H., (1990) A new therapeutic agent: azelaic acid in acne treatment. *J Dermatolog Treat* **1**: 23-28.

Gollub, J., C. A. Ball, G. Binkley, J. Demeter, D. B. Finkelstein, J. M. Hebert, T. Hernandez-Boussard, H. Jin, M. Kaloper, J. C. Matese, M. Schroeder, P. O. Brown, D. Botstein & G. Sherlock, (2003) The Stanford Microarray Database: data access and quality assessment tools. *Nucleic Acids Res* **31**: 94-96.

Graham, G. M., M. D. Farrar, J. E. Cruse-Sawyer, K. T. Holland & E. Ingham, (2004) Proinflammatory cytokine production by human keratinocytes stimulated with *Propionibacterium acnes* and *P. acnes* GroEL. *Br J Dermatol* **150**: 421-428.

Greenman, J. & K. Holland, (1985) Effects of dilution rate on biomass and extracellular enzyme production by three species of cutaneous propionibacteria grown in continuous culture. *J Gen Microbiol* **131**: 1619-1624.

Greenman, J., K. T. Holland & W. J. Cunliffe, (1981) Effects of glucose concentration n biomass, maximum specific growth rate and extracellular enzyme production by three species of cutaneous propionibacteria grown in continuous culture. *J Gen Microbiol* **127**: 371-376.

Greenman, J., K. T. Holland & W. J. Cunliffe, (1983) Effects of pH on biomass, maximum specific growth rate and extracellular enzyme production by three species of cutaneous propionibacteria grown in continuous culture. *J Gen Microbiol* **129**: 1301-1307.

Grice, E. A., H. H. Kong, G. Renaud, A. C. Young, G. G. Bouffard, R. W. Blakesley, T. G. Wolfsberg, M. L. Turner & J. A. Segre, (2008) A diversity profile of the human skin microbiota. *Genome Res* **18**: 1043-1050.

Grice, E. A. & J. A. Segre, (2011) The skin microbiome. *Nat Rev Microbiol* **9**: 244-253.

Griffiths-Jones, S., S. Moxon, M. Marshall, A. Khanna, S. R. Eddy & A. Bateman, (2005) Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* **33**: D121-D124.

Guan, S., (2011) A novel two-component signal transduction system in *Propionibacterium acnes* and its association with a putative extracellular signalling peptide. In.: University of Leeds, pp.

Guell, M., V. van Noort, E. Yus, W. H. Chen, J. Leigh-Bell, K. Michalodimitrakis, T. Yamada, M. Arumugam, T. Doerks, S. Kuhner, M. Rode, M. Suyama, S. Schmidt, A. C. Gavin, P. Bork & L. Serrano, (2009) Transcriptome complexity in a genome-reduced bacterium. *Science* **326**: 1268-1271.

Gust, B., G. Chandra, D. Jakimowicz, T. Yuqing, C. J. Bruton & K. F. Chater, (2004) λ Red-mediated genetic manipulation of antibiotic-producing *Streptomyces*. *Adv Appl Microbiol* **54**: 107-128.

Hall-Stoodley, L., J. W. Costerton & P. Stoodley, (2004) Bacterial biofilms: from the natural environment to infectious diseases. *Nat Rev Microbiol* **2**: 95-108.

Hamann, K., P. Zimmann & K. Altendorf, (2008) Reduction of turgor is not the stimulus for the sensor kinase KdpD of *Escherichia coli*. *J Bacteriol* **190**: 2360-2367.

Harder, W. & J. G. Kuenen, (1977) A review. Microbial selection in continuous culture. *J Appl Bacteriol* **43**: 1-24.

Harley, C. B. & R. P. Reynolds, (1987) Analysis of *Escherichia coli* promoter sequences. *Nucleic Acids Res.* **15**: 2343-2361.

Hartmann, R. K., M. Gossringer, B. Spath, S. Fischer & A. Marchfelder, (2009) The Making of tRNAs and More - RNase P and tRNase Z. *Molecular Biology of RNA Processing and Decay in Prokaryotes* **85**: 319-368.

Hashimshony, T., F. Wagner, N. Sher & I. Yanai, (2012) CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep* **2**: 666-673.

Hassing, G. S., (1971) Partial purification and some properties of a lipase from *Corynebacterium acnes*. *Biochimica et Biophysica Acta (BBA)-Enzymology* **242**: 381-394.

He, S., O. Wurtzel, K. Singh, J. L. Froula, S. Yilmaz, S. G. Tringe, Z. Wang, F. Chen, E. A. Lindquist & R. Sorek, (2010) Validation of two ribosomal RNA removal methods for microbial metatranscriptomics. *Nature methods* **7**: 807-812.

Henkin, T. M. & C. Yanofsky, (2002) Regulation by transcription attenuation in bacteria: how RNA provides instructions for transcription termination/antitermination decisions. *BioEssays* **24**: 700-707.

Hiard, S., R. Maree, S. Colson, P. A. Hoskisson, F. Titgemeyer, G. P. van Wezel, B. Joris, L. Wehenkel & S. Rigali, (2007) PREDetector: A new tool to identify regulatory elements in bacterial genomes. *Biochem. Biophys. Res. Commun.* **357**: 861-864.

Hoch, J. A., (2000) Two-component and phosphorelay signal transduction. *Curr Opin Microbiol* **3**: 165-170.

Hoch, J. A., T. J. Silhavy & R. B. Bourret, (1995) *Two-component signal transduction*. ASM press Washington, DC.

Holland, D. B., R. A. Bojar, M. D. Farrar & K. T. Holland, (2009) Differential innate immune responses of a living skin equivalent model colonized by *Staphylococcus epidermidis* or *Staphylococcus aureus*. *FEMS Microbiol Lett* **290**: 149-155.

Holland, D. B., R. A. Bojar, A. H. Jeremy, E. Ingham & K. T. Holland, (2008) Microbial colonization of an in vitro model of a tissue engineered human skin equivalent - a novel approach. *FEMS Microbiol Lett* **279**: 110-115.

Holland, D. B., W. J. Cunliffe & J. F. Norris, (1998) Differential response of sebaceous glands to exogenous testosterone. *Br J Dermatol* **139**: 102-103.

Holland, K. & R. Bojar, (1993) Antimicrobial effects of azelaic acid. *J Dermatolog Treat* **4**: 8-11.

Holland, K. T., W. J. Cunliffe & C. D. Roberts, (1978) The role of bacteria in acne vulgaris: a new approach. *Clin Exp Dermatol* **3**: 253-257.

Holland, K. T., J. Greenman & W. J. Cunliffe, (1979) Growth of cutaneous propionibacteria on synthetic medium; growth yields and exoenzyme production. *J Appl Bacteriol* **47**: 383-394.

Holland, K. T., E. Ingham & W. J. Cunliffe, (1981) A review, the microbiology of acne. *J Appl Bacteriol* **51**: 195-215.

Holmberg, A., R. Lood, M. Morgelin, B. Soderquist, E. Holst, M. Collin, B. Christensson & M. Rasmussen, (2009) Biofilm formation by *Propionibacterium acnes* is a characteristic of invasive isolates. *Clin Microbiol Infect* **15**: 787-795.

Horváth, B., J. Hunyadkürti, A. Vörös, C. Fekete, E. Urbán, L. Kemény & I. Nagy, (2012) Genome sequence of *Propionibacterium acnes* type II strain ATCC 11828. *J Bacteriol* **194**: 202-203.

Hovatta, I., K. Kimppa, A. Lehmussola, T. Pasanen, J. Saarela, I. Saarikko, J. Saharinen, P. Tiikkainen, T. Toivanen, M. Tolvanen, M. Vihinen & G. Wong, (2005) *DNA Microarray Data Analysis*, p. 163. CSC - Scientific Computing Ltd, Helsinki.

Hsing, W., F. D. Russo, K. K. Bernd & T. J. Silhavy, (1998) Mutations that alter the kinase and phosphatase activities of the two-component sensor EnvZ. *J Bacteriol* **180**: 4538-4546.

Huerta, A. M. & J. Collado-Vides, (2003) Sigma70 promoters in *Escherichia coli*: specific transcription in dense regions of overlapping promoter-like signals. *J Mol Biol* **333**: 261-278.

Hunyadkürti, J., Z. Feltóti, B. Horváth, M. Nagymihály, A. Vörös, A. McDowell, S. Patrick, E. Urbán & I. Nagy, (2011) Complete genome sequence of *Propionibacterium acnes* Type IB Strain 6609. *J Bacteriol* **193**: 4561-4562.

Hutchison, C. A., 3rd, H. O. Smith, C. Pfannkoch & J. C. Venter, (2005) Cell-free cloning using ϕ29 DNA polymerase. *Proc Natl Acad Sci USA* **102**: 17332-17336.

Ingham, E., E. A. Eady, C. E. Goodwin, J. H. Cove & W. J. Cunliffe, (1992) Pro-inflammatory levels of interleukin-1α-like bioactivity are present in the majority of open comedones in acne vulgaris. *J Invest Dermatol* **98**: 895-901.

Ingham, E., K. T. Holland, G. Gowland & W. J. Cunliffe, (1979) Purification and partial characterization of hyaluronate lyase (EC 4.2.2.1) from *Propionibacterium acnes*. *J Gen Microbiol* **115**: 411-418.

Ingham, E., K. T. Holland, G. Gowland & W. J. Cunliffe, (1980) Purification and partial characterization of an acid phosphatase (EC 3.1.3.2) produced by *Propionibacterium acnes*. *J Gen Microbiol* **118**: 59-65.

Ingham, E., K. T. Holland, G. Gowland & W. J. Cunliffe, (1981) Partial purification and characterization of lipase (EC 3.1.1.3) from *Propionibacterium acnes*. *J Gen Microbiol* **124**: 393-401.

Ingham, E., C. Walters, E. Eady, J. Cove, J. Kearney & W. Cunliffe, (1998) Inflammation in acne vulgaris: failure of skin micro-organisms to modulate keratinocyte interleukin 1α production *in vitro*. *Dermatology* **196**: 86-88.

Ingram, E., K. Holland, G. Gowland & W. Cunliffe, (1983) Studies of the extracellular proteolytic activity produced by *Propionibacterium acnes*. *J Appl Microbiol* **54**: 263-271.

Irizarry, R. A., B. M. Bolstad, F. Collin, L. M. Cope, B. Hobbs & T. P. Speed, (2003) Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* **31**: e15.

Jacobs, D. G., N. L. Deutsch & M. Brewer, (2001) Suicide, depression, and isotretinoin: is there a causal link? *J Am Acad Dermatol* **45**: S168-175.

Jacquier, A., (2009) The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. *Nat. Rev. Genet.* **10**: 833-844.

Jager, D., C. M. Sharma, J. Thomsen, C. Ehlers, J. Vogel & R. A. Schmitz, (2009) Deep sequencing analysis of the *Methanosarcina mazei* Go1 transcriptome in response to nitrogen availability. *Proc. Natl. Acad. Sci. USA* **106**: 21878-21882.

Jahns, A. C., B. Lundskog, R. Ganceviciene, R. H. Palmer, I. Golovleva, C. C. Zouboulis, A. McDowell, S. Patrick & O. A. Alexeyev, (2012) An increased incidence of *Propionibacterium acnes* biofilms in acne vulgaris: a case-control study. *Br J Dermatol* **167**: 50-58.

Jakab, E., R. Zbinden, J. Gubler, C. Ruef, A. von Graevenitz & M. Krause, (1996) Severe infections caused by *Propionibacterium acnes*: an underestimated pathogen in late postoperative infections. *Yale J Biol Med* **69**: 477-482.

Janssen, G. R., (1993) Eubacterial, archaebacterial, and eukaryotic genes that encode leaderless mRNA. *Industrial Microorganisms: Basic and Applied Molecular Genetics*: 59-67.

Jappe, U., (2003) Pathological mechanisms of acne with special emphasis on *Propionibacterium acnes* and related therapy. *Acta Derm Venereol* **83**: 241-248.

Jarrige, A. C., N. Mathy & C. Portier, (2001) PNPase autocontrols its expression by degrading a double-stranded structure in the *pnp* mRNA leader. *EMBO J.* **20**: 6845-6855.

Jayapal, K. P., R. J. Philp, Y. J. Kok, M. G. Yap, D. H. Sherman, T. J. Griffin & W. S. Hu, (2008) Uncovering genes with divergent mRNA-protein dynamics in *Streptomyces coelicolor*. *PLoS One* **3**: e2097.

Jung, K., K. Hamann & A. Revermann, (2001) K^+ Stimulates Specifically the Autokinase Activity of Purified and Reconstituted EnvZ of *Escherichia coli*. *J. Biol. Chem.* **276**: 40896-40902.

Kafatos, F. C., C. W. Jones & A. Efstratiadis, (1979) Determination of nucleic acid sequence homologies and relative concentrations by a dot hybridization procedure. *Nucleic Acids Res* **7**: 1541-1552.

Kanaar, P., (1971) Follicular-keratogenic properties of fatty acids in the external ear canal of the rabbit. *Dermatologica* **142**: 14-22.

Kanehisa, M., S. Goto, Y. Sato, M. Furumichi & M. Tanabe, (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* **40**: D109-114.

Kang, Y., M. H. Norris, J. Zarzycki-Siek, W. C. Nierman, S. P. Donachie & T. T. Hoang, (2011) Transcript amplification from single bacterium for transcriptome analysis. *Genome Res* **21**: 925-935.

Karr, J. R., J. C. Sanghvi, D. N. Macklin, A. Arora & M. W. Covert, (2012a) WholeCellKB: model organism databases for comprehensive whole-cell models. *Nucleic Acids Res.*

Karr, J. R., J. C. Sanghvi, D. N. Macklin, M. V. Gutschow, J. M. Jacobs, B. Bolival, N. Assad-Garcia, J. I. Glass & M. W. Covert, (2012b) A whole-cell computational model predicts phenotype from genotype. *Cell* **150**: 389-401.

Kieser, T., M. J. Bibb, M. J. Buttner, K. F. Chater & D. A. Hopwood, (2000) *Practical streptomyces genetics*. John Innes Foundation Norwich, UK.

Kim, J., (2005) Review of the innate immune response in acne vulgaris: activation of Toll-like receptor 2 in acne triggers inflammatory cytokine responses. *Dermatology* **211**: 193-198.

Kim, J., M. T. Ochoa, S. R. Krutzik, O. Takeuchi, S. Uematsu, A. J. Legaspi, H. D. Brightbill, D. Holland, W. J. Cunliffe, S. Akira, P. A. Sieling, P. J. Godowski & R. L. Modlin, (2002) Activation of toll-like receptor 2 in acne triggers inflammatory cytokine responses. *J Immunol* **169**: 1535-1541.

Kligman, A. M., (1968) Pathogenesis of acne vulgaris. II. Histopathology of comedones induced in the rabbit ear by human sebum. *Arch Dermatol* **98**: 58-66.

Kligman, A. M., (1974) An overview of acne. *J Invest Dermatol* **62**: 268-287.

Koo, J. Y. & L. L. Smith, (1991) Psychologic aspects of acne. *Pediatr Dermatol* **8**: 185-188.

Kornberg, A., N. N. Rao & D. Ault-Riche, (1999) Inorganic polyphosphate: a molecule of many functions. *Annu Rev Biochem* **68**: 89-125.

Laing, E. & C. P. Smith, (2010) RankProdIt: A web-interactive Rank Products analysis tool. *BMC Res Notes* **3**: 221.

Langmead, B. & S. L. Salzberg, (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**: 357-U354.

Lasa, I., A. Toledo-Arana, A. Dobin, M. Villanueva, I. R. de los Mozos, M. Vergara-Irigaray, V. Segura, D. Fagegaltier, J. R. Penades, J. Valle, C. Solano & T. R. Gingeras, (2011) Genome-wide antisense transcription drives mRNA processing in bacteria. *Proc. Natl. Acad. Sci. USA* **108**: 20172-20177.

Law, M. P., A. A. Chuh, A. Lee & N. Molinari, (2010) Acne prevalence and beyond: acne disability and its predictive factors among Chinese late adolescents in Hong Kong. *Clin Exp Dermatol* **35**: 16-21.

Lee, S. Y., S. C. Bailey & D. Apirion, (1978) Small stable RNAs from *Escherichia coli*: evidence for existence of new molecules and for a new ribonucleoprotein particle containing 6S RNA. *J. Bacteriol.* **133**: 1015-1023.

Lee, W. J., H. D. Jung, H. J. Lee, B. S. Kim, S.-J. Lee & D. W. Kim, (2008) Influence of substance-P on cultured sebocytes. *Arch Dermatol Res* **300**: 311-316.

Levin, J. Z., M. Yassour, X. Adiconis, C. Nusbaum, D. A. Thompson, N. Friedman, A. Gnirke & A. Regev, (2010) Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nat Methods* **7**: 709-715.

Levy, P. Y., F. Fenollar, A. Stein, F. Borrione, E. Cohen, B. Lebail & D. Raoult, (2008) *Propionibacterium acnes* postoperative shoulder arthritis: an emerging clinical entity. *Clin Infect Dis* **46**: 1884-1886.

Lin-Chao, S. & S. N. Cohen, (1991) The rate of processing and degradation of antisense RNAI regulates the replication of ColE1-type plasmids *in vivo*. *Cell* **65**: 1233-1242.

Lin, M. H., J. C. Shu, H. Y. Huang & Y. C. Cheng, (2012) Involvement of iron in biofilm formation by *Staphylococcus aureus*. *PLoS One* **7**: e34388.

Lindsay, D. & A. von Holy, (2006) Bacterial biofilms within the clinical setting: what healthcare professionals should know. *J Hosp Infect* **64**: 313-325.

Lisser, S. & H. Margalit, (1993) Compilation of *Escherichia coli* mRNA promoter sequences. *Nucleic Acids Res.* **21**: 1507-1516.

Liu, J. M., J. Livny, M. S. Lawrence, M. D. Kimball, M. K. Waldor & A. Camilli, (2009) Experimental discovery of sRNAs in *Vibrio cholerae* by direct cloning, 5S/tRNA depletion and parallel sequencing. *Nucleic Acids Res.* **37**.

Lyons, R. E., (1978) Comparative effectiveness of benzoyl peroxide and tretinoin in acne vulgaris. *Int J Dermatol* **17**: 246-251.

Mackenzie, S., P. Abraham, J. Broom, M. Morris, H. R. Crocker, L. Roberts, J. F. Payne, H. Waldo & G. S. Taylor, (1894) A discussion on the etiology and treatment of acne vulgaris. *Brit Med J* **2**: 688-692.

Malys, N. & J. E. G. McCarthy, (2011) Translation initiation: variations in the mechanism can be anticipated. *Cell. Mol. Life Sci.* **68**: 991-1003.

Mamanova, L., R. M. Andrews, K. D. James, E. M. Sheridan, P. D. Ellis, C. F. Langford, T. W. Ost, J. E. Collins & D. J. Turner, (2010a) FRT-seq: amplification-free, strand-specific transcriptome sequencing. *Nature methods* **7**: 130-132.

Mamanova, L., R. M. Andrews, K. D. James, E. M. Sheridan, P. D. Ellis, C. F. Langford, T. W. B. Ost, J. E. Collins & D. J. Turner, (2010b) FRT-seq: amplification-free, strand-specific transcriptome sequencing. *Nat. Methods* **7**: 130-U163.

Mamanova, L. & D. J. Turner, (2011) Low-bias, strand-specific transcriptome Illumina sequencing by on-flowcell reverse transcription (FRT-seq). *Nat Protoc* **6**: 1736-1747.

Marguerat, S. & J. Bahler, (2010) RNA-seq: from technology to biology. *Cell. Mol. Life Sci.* **67**: 569-579.

Marincs, F., I. W. Manfield, J. A. Stead, K. J. McDowall & P. G. Stockley, (2006) Transcript analysis reveals an extended regulon and the importance of protein-protein co-operativity for the *Escherichia coli* methionine repressor. *Biochem. J.* **396**: 227-234.

Marioni, J. C., C. E. Mason, S. M. Mane, M. Stephens & Y. Gilad, (2008) RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res* **18**: 1509-1517.

Martin, J., W. Zhu, K. D. Passalacqua, N. Bergman & M. Borodovsky, (2010a) *Bacillus anthracis* genome organization in light of whole transcriptome sequencing. *BMC Bioinformatics* **11 Suppl 3**: S10.

Martin, J., W. H. Zhu, K. D. Passalacqua, N. Bergman & M. Borodovsky, (2010b) *Bacillus anthracis* genome organization in light of whole transcriptome sequencing. *Bmc Bioinformatics* **11**.

Marynick, S. P., Z. H. Chakmakjian, D. L. McCaffree & J. H. Herndon, Jr., (1983) Androgen excess in cystic acne. *N Engl J Med* **308**: 981-986.

Masse, E., F. E. Escorcia & S. Gottesman, (2003) Coupled degradation of a small regulatory RNA and its mRNA targets in *Escherichia coli*. *Genes Dev* **17**: 2374-2383.

Mathy, N., L. Benard, O. Pellegrini, R. Daou, T. Y. Wen & C. Condon, (2007) 5 '-to-3 ' exoribonuclease activity in bacteria: Role of RNase J1 in rRNA maturation and 5 ' stability of mRNA. *Cell* **129**: 681-692.

Matsushima, P., M. C. Broughton, J. R. Turner & R. H. Baltz, (1994) Conjugal transfer of cosmid DNA from *Escherichia coli* to *Saccharopolyspora spinosa*: effects of chromosomal insertions on macrolide A83543 production. *Gene* **146**: 39-45.

Mazurkiewicz, P., C. M. Tang, C. Boone & D. W. Holden, (2006) Signature-tagged mutagenesis: barcoding mutants for genome-wide screens. *Nat Rev Genet* **7**: 929-939.

McLorinan, G. C., J. V. Glenn, M. G. McMullan & S. Patrick, (2005) *Propionibacterium acnes* wound contamination at the time of spinal surgery. *Clin Orthop Relat Res*: 67-73.

Mendoza-Vargas, A., L. Olvera, M. Olvera, R. Grande, L. Vega-Alvarado, B. Taboada, V. Jimenez-Jacinto, H. Salgado, K. Juarez, B. Contreras-Moreira, A. M. Huerta, J. Collado-Vides & E. Morett, (2009) Genome-wide identification of transcription start sites, promoters and transcription factor binding sites in *E. coli*. *Plos One* **4**.

Merritt, J., F. Qi, S. D. Goodman, M. H. Anderson & W. Shi, (2003) Mutation of *luxS* affects biofilm formation in *Streptococcus mutans*. *Infect Immun* **71**: 1972-1979.

Michoel, T., R. De Smet, A. Joshi, K. Marchal & Y. Van de Peer, (2009) Reverse-engineering transcriptional modules from gene expression data. *Ann N Y Acad Sci* **1158**: 36-43.

Mitschke, J., J. Georg, I. Scholz, C. M. Sharma, D. Dienst, J. Bantscheff, B. Voss, C. Steglich, A. Wilde, J. Vogel & W. R. Hess, (2011) An experimentally anchored map of transcriptional start sites in the model cyanobacterium *Synechocystis* sp PCC6803. *Proc. Natl. Acad. Sci. USA* **108**: 2124-2129.

Mohanty, B. K. & S. R. Kushner, (2000) Polynucleotide phosphorylase functions both as a 3 '-> 5 ' exonuclease and a poly(A) polymerase in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **97**: 11966-11971.

Moll, I., S. Grill, C. O. Gualerzi & U. Blasi, (2002) Leaderless mRNAs in bacteria: surprises in ribosomal recruitment and translational control. *Mol. Microbiol.* **43**: 239-246.

Molle, V., M. Fujita, S. T. Jensen, P. Eichenberger, J. E. Gonzalez-Pastor, J. S. Liu & R. Losick, (2003) The Spo0A regulon of *Bacillus subtilis*. *Mol Microbiol* **50**: 1683-1701.

Monod, J., (1949) The growth of bacterial cultures. *Annu Rev Microbiol* **3**: 371-394.

Motoyoshi, K., (1983) Enhanced comedo formation in rabbit ear skin by squalene and oleic acid peroxides. *Br J Dermatol* **109**: 191-198.

Nahvi, A., N. Sudarsan, M. S. Ebert, X. Zou, K. L. Brown & R. R. Breaker, (2002) Genetic control by a metabolite binding mRNA. *Chem. Biol.* **9**: 1043-1049.

Neidhardt, F. C., J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter & H. Umbarger, (1987) *Escherichia coli and Salmonella typhimurium. Cellular and molecular biology. Volumes I and II*. American Society for Microbiology.

Nguyen, R. & J. Su, (2011) Treatment of acne vulgaris. *Paediatrics and Child Health* **21**: 119-125.

Nicholson, A. W., (2003) The ribonuclease III superfamily: forms and functions in RNA maturation, decay, and gene silencing In: RNAi: A Guide to Gene Silencing. G. J. Hannon (ed). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, pp.

Nisbet, M., S. Briggs, R. Ellis-Pegler, M. Thomas & D. Holland, (2007) *Propionibacterium acnes*: an under-appreciated cause of post-neurosurgical infection. *J Antimicrob Chemother* **60**: 1097-1103.

Noble, W. C., (1984) Skin microbiology: coming of age. *J Med Microbiol* **17**: 1-12.

Nogueira, T. & M. Springer, (2000) Post-transcriptional control by global regulators of gene expression in bacteria. *Curr Opin Microbiol* **3**: 154-158.

Nordlund, N. & P. Reichard, (2006) Ribonucleotide reductases. *Annu. Rev. Biochem.* **75**: 681-706.

Novick, A., (1955) Growth of bacteria. *Annu Rev Microbiol* **9**: 97-110.

O'Donnell, S. A. & G. R. Janssen, (2002) Leaderless mRNAs bind 70S ribosomes more strongly than 30S ribosomal subunits in *Escherichia coli*. *J. Bacteriol.* **184**: 6730-6733.

Oliver, J. D., (2005) The viable but nonculturable state in bacteria. *J Microbiol* **43 Spec No**: 93-100.

Ostlere, L. S., G. Rumsby, P. Holownia, H. S. Jacobs, M. H. Rustin & J. W. Honour, (1998) Carrier status for steroid 21-hydroxylase deficiency is only one factor in the variable phenotype of acne. *Clin Endocrinol* **48**: 209-215.

Papageorgiou, P., A. Katsambas & A. Chu, (2000) Phototherapy with blue (415 nm) and red (660 nm) light in the treatment of acne vulgaris. *Br J Dermatol* **142**: 973-978.

Parkinson, J. S., (1993) Signal transduction schemes of bacteria. *Cell* **73**: 857-871.

Pochi, P. E., J. S. Strauss & D. T. Downing, (1979) Age-related changes in sebaceous gland activity. *J Invest Dermatol* **73**: 108-111.

Polarek, J. W., G. Williams & W. Epstein, (1992) The products of the *kdpDE* operon are required for expression of the Kdp ATPase of *Escherichia coli*. *J Bacteriol* **174**: 2145-2151.

Pruitt, K. D., T. Tatusova & D. R. Maglott, (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **35**: D61-D65.

Pushkarev, D., N. F. Neff & S. R. Quake, (2009) Single-molecule sequencing of an individual human genome. *Nat Biotechnol* **27**: 847-850.

Qin, J., Y. Li, Z. Cai, S. Li, J. Zhu, F. Zhang, S. Liang, W. Zhang, Y. Guan, D. Shen, Y. Peng, D. Zhang, Z. Jie, W. Wu, Y. Qin, W. Xue, J. Li, L. Han, D. Lu, P. Wu, Y. Dai, X. Sun, Z. Li, A. Tang, S. Zhong, X. Li, W. Chen, R. Xu, M. Wang, Q. Feng, M. Gong, J. Yu, Y. Zhang, M. Zhang, T. Hansen, G. Sanchez, J. Raes, G. Falony, S. Okuda, M. Almeida, E. LeChatelier, P. Renault, N. Pons, J. M. Batto, Z. Zhang, H. Chen, R. Yang, W. Zheng, H. Yang, J. Wang, S. D. Ehrlich,

R. Nielsen, O. Pedersen & K. Kristiansen, (2012) A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490**: 55-60.

Quinlan, A. R. & I. M. Hall, (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841-842.

Rademaker, M., (2012) Isotretinoin: dose, duration and relapse. What does 30 years of usage tell us? *Australas J Dermatol.*

Radtke, M. A., I. Schafer & M. Augustin, (2010) Pharmacoeconomy in acne - evaluation of benefit and economics. *J Dtsch Dermatol Ges* **8 Suppl 1**: S105-114.

Ramage, G., M. M. Tunney, S. Patrick, S. P. Gorman & J. R. Nixon, (2003) Formation of *Propionibacterium acnes* biofilms on orthopaedic biomaterials and their susceptibility to antimicrobials. *Biomaterials* **24**: 3221-3227.

Rasmussen, S., H. B. Nielsen & H. Jarmer, (2009) The transcriptionally active regions in the genome of *Bacillus subtilis*. *Mol. Microbiol.* **73**: 1043-1057.

Resch, A., R. Rosenstein, C. Nerz & F. Gotz, (2005) Differential gene expression profiling of *Staphylococcus aureus* cultivated under biofilm and planktonic conditions. *Appl Environ Microbiol* **71**: 2663-2676.

Robert-Le meur, M. & C. Portier, (1992) *Escherichia coli* polynucleotide phosphorylase expression is autoregulated through an RNase III-dependent mechanism. *EMBO J.* **11**: 2633-2641.

Robert-Le meur, M. & C. Portier, (1994) Polynucleotide phosphorylase of *Escherichia coli* induces the degradation of its RNase III-processed messenger by preventing its translation. *Nucleic Acids Res.* **22**: 397-403.

Robertson, H. D., R. E. Webster & N. D. Zinder, (1968) Purification and properties of ribonuclease III from *Escherichia coli*. *J. Biol. Chem.* **243**: 82-&.

Ross, J., A. Snelling, E. Carnegie, P. Coates, W. Cunliffe, V. Bettoli, G. Tosti, A. Katsambas, J. Galvan Perez Del Pulgar & O. Rollman, (2003) Antibiotic-resistant acne: lessons from Europe. *Br J Dermatol* **148**: 467-478.

Ryder, V. J., (2010) The mutability of staphyloccal biofilms. In.: University of Leeds, pp.

Salgado, H., S. Gama-Castro, M. Peralta-Gil, E. Diaz-Peredo, F. Sanchez-Solano, A. Santos-Zavaleta, I. Martinez-Flores, V. Jimenez-Jacinto, C. Bonavides-Martinez, J. Segura-Salazar, A. Martinez-Antonio & J. Collado-Vides, (2006) RegulonDB (version 5.0): *Escherichia coli* K-12 transcriptional regulatory network, operon organization, and growth conditions. *Nucleic Acids Res.* **34**: D394-D397.

Salgado, H., M. Peralta-Gil, S. Gama-Castro, A. Santos-Zavaleta, L. Muniz-Rascado, J. S. Garcia-Sotelo, V. Weiss, H. Solano-Lira, I. Martinez-Flores, A. Medina-Rivera, G. Salgado-Osorio, S. Alquicira-Hernandez, K. Alquicira-Hernandez, A. Lopez-Fuentes, L. Porron-Sotelo, A. M. Huerta, C. Bonavides-Martinez, Y. I. Balderas-Martinez, L. Pannier, M. Olvera, A. Labastida, V. Jimenez-Jacinto, L. Vega-Alvarado, V. Del Moral-Chavez, A. Hernandez-Alvarez, E. Morett & J. Collado-Vides, (2013) RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. *Nucleic Acids Res* **41**: D203-213.

Sambrook, J. & D. W. Russell, (2006) Agarose gel electrophoresis. *CSH Protoc* **2006**: pdb. prot4020.

Schena, M., D. Shalon, R. W. Davis & P. O. Brown, (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**: 467-470.

Schneider, K. L., K. S. Pollard, R. Baertsch, A. Pohl & T. M. Lowe, (2006) The UCSC Archaeal Genome Browser. *Nucleic Acids Res* **34**: D407-410.

Shahbabian, K., A. Jamalli, L. Zig & H. Putzer, (2009) RNase Y, a novel endoribonuclease, initiates riboswitch turnover in *Bacillus subtilis*. *EMBO J.* **28**: 3523-3533.

Sharma, C. M., S. Hoffmann, F. Darfeuille, J. Reignier, S. Findeiss, A. Sittka, S. Chabas, K. Reiche, J. Hackermuller, R. Reinhardt, P. F. Stadler & J. Vogel, (2010) The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature* **464**: 250-255.

Shelness, G. S. & D. L. Williams, (1985) Secondary structure analysis of apolipoprotein II mRNA using enzymatic probes and reverse transcriptase. Evaluation of primer extension for high resolution structure mapping of mRNA. *J Biol Chem* **260**: 8637-8646.

Shine, J. & L. Dalgarno, (1974) 3'-terminal sequence of *Escherichia coli* 16S rRNA: possible role in initiation and termination of protein synthesis. *P. Aust. Biochem. Soc.* **7**: 72-72.

Shine, J. & L. Dalgarno, (1975) Terminal sequence analysis of bacterial rRNA: correlation between 3'-terminal polypyrimidine sequence of 16S RNA and translational specificity of ribosome. *Eur. J. Biochem.* **57**: 221-230.

Simonart, T., M. Dramaix & V. De Maertelaer, (2008) Efficacy of tetracyclines in the treatment of acne vulgaris: a review. *Br J Dermatol* **158**: 208-216.

Simpson, N., (1993) Social and economic aspects of acne and the cost-effectiveness of isotretinoin. *J Dermatolog Treat* **4**: 6-9.

Sörensen, M., T. N. Mak, R. Hurwitz, L. A. Ogilvie, H. J. Mollenkopf, T. F. Meyer & H. Brüggemann, (2010) Mutagenesis of *Propionibacterium acnes* and analysis of two CAMP factor knock-out mutants. *J Microbiol Methods* **83**: 211-216.

Stepanovic, S., D. Vukovic, I. Dakic, B. Savic & M. Svabic-Vlahovic, (2000) A modified microtiter-plate test for quantification of staphylococcal biofilm formation. *J Microbiol Methods* **40**: 175-179.

Stern, R. S., (2000) Medication and medical service utilization for acne 1995-1998. *J Am Acad Dermatol* **43**: 1042-1048.

Stewart, G. C. & K. F. Bott, (1983) DNA sequence of the tandem rRNA promoter for *B subtilis* operon *rrnB*. *Nucleic Acids Res.* **11**: 6289-6300.

Stewart, M. E., R. Greenwood, W. J. Cunliffe, J. S. Strauss & D. T. Downing, (1986) Effect of cyproterone acetate-ethinyl estradiol treatment on the proportions of linoleic and sebaleic acids in various skin surface lipid classes. *Arch Dermatol Res* **278**: 481-485.

Stewart, P. S. & M. J. Franklin, (2008) Physiological heterogeneity in biofilms. *Nat Rev Microbiol* **6**: 199-210.

Stock, A. M., V. L. Robinson & P. N. Goudreau, (2000) Two-component signal transduction. *Annu Rev Biochem* **69**: 183-215.

Stoll, S., A. R. Shalita, G. F. Webster, R. Kaplan, S. Danesh & A. Penstein, (2001) The effect of the menstrual cycle on acne. *J Am Acad Dermatol* **45**: 957-960.

Stoodley, P., D. Debeer & Z. Lewandowski, (1994) Liquid flow in biofilm systems. *Appl Environ Microbiol* **60**: 2711-2716.

Storz, G., J. A. Opdyke & A. Zhang, (2004) Controlling mRNA stability and translation with small, noncoding RNAs. *Curr Opin Microbiol* **7**: 140-144.

Suelter, C. H., (1970) Enzymes activated by monovalent cations. *Science* **168**: 789-795.

Sugiura, A., K. Nakashima & T. Mizuno, (1993) Sequence-directed DNA curvature in activator-binding sequence in the *Escherichia coli kdpABC* promoter. *Biosci Biotechnol Biochem* **57**: 356-357.

Sugiura, A., K. Nakashima, K. Tanaka & T. Mizuno, (1992) Clarification of the structural and functional features of the osmoregulated *kdp* operon of *Escherichia coli*. *Mol Microbiol* **6**: 1769-1776.

Suzek, B. E., M. D. Ermolaeva, M. Schreiber & S. L. Salzberg, (2001) A probabilistic, method for identifying start codons in bacterial genomes. *Bioinformatics* **17**: 1123-1130.

Tafin, U. F., S. Corvec, B. Betrisey, W. Zimmerli & A. Trampuz, (2012) Role of rifampin against *Propionibacterium acnes* biofilm in vitro and in an experimental foreign-body infection model. *Antimicrob Agents Chemother* **56**: 1885-1891.

Tang, F., C. Barbacioru, Y. Wang, E. Nordman, C. Lee, N. Xu, X. Wang, J. Bodeau, B. B. Tuch, A. Siddiqui, K. Lao & M. A. Surani, (2009) mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* **6**: 377-382.

Thomas, L., D. A. Hodgson, A. Wentzel, K. Nieselt, T. E. Ellingsen, J. Moore, E. R. Morrissey, R. Legaie, W. Wohlleben, A. Rodriguez-Garcia, J. F. Martin, N. J. Burroughs, E. M. Wellington & M. C. Smith, (2012) Metabolic switches and adaptations deduced from the proteomes of *Streptomyces coelicolor* wild type and *phoP* mutant grown in batch culture. *Mol Cell Proteomics* **11**: M111 013797.

Timmermans, J. & L. Van Melderen, (2010) Post-transcriptional global regulation by CsrA in bacteria. *Cell Mol Life Sci* **67**: 2897-2908.

Toledo-Arana, A., O. Dussurget, G. Nikitas, N. Sesto, H. Guet-Revillet, D. Balestrino, E. Loh, J. Gripenland, T. Tiensuu, K. Vaitkevicius, M. Barthelemy, M. Vergassola, M. A. Nahori, G. Soubigou, B. Regnault, J. Y. Coppee, M. Lecuit, J. Johansson & P. Cossart, (2009) The *Listeria* transcriptional landscape from saprophytism to virulence. *Nature* **459**: 950-956.

Torrents, E., I. Grinberg, B. Gorovitz-Harris, H. Lundstrom, I. Borovok, Y. Aharonowitz, B. M. Sjoberg & G. Cohen, (2007) NrdR controls differential expression of the *Escherichia coli* ribonucleotide reductase genes. *J. Bacteriol.* **189**: 5012-5021.

Toyoda, M. & M. Morohashi, (2001) Pathogenesis of acne. *Medical Electron Microscopy* **34**: 29-40.

Tunney, M. M., S. Patrick, M. D. Curran, G. Ramage, D. Hanna, J. R. Nixon, S. P. Gorman, R. I. Davis & N. Anderson, (1999) Detection of prosthetic hip infection at revision arthroplasty by immunofluorescence microscopy and PCR amplification of the bacterial 16S rRNA gene. *J Clin Microbiol* **37**: 3281-3290.

Van Dessel, W., L. Van Mellaert, N. Geukens, E. Lammertyn & J. Anne, (2004) Isolation of high quality RNA from *Streptomyces*. *J Microbiol Methods* **58**: 135-137.

Van Etten, W. J. & G. R. Janssen, (1998) An AUG initiation codon, not codon-anticodon complementarity, is required for the translation of unleadered mRNA in *Escherichia coli*. *Mol. Microbiol.* **27**: 987-1001.

Vesper, O., S. Amitai, M. Belitsky, K. Byrgazov, A. C. Kaberdina, H. Engelberg-Kulka & I. Moll, (2011) Selective translation of leaderless mRNAs by specialized ribosomes generated by MazF in *Escherichia coli*. *Cell* **147**: 147-157.

Vioque, A., J. Arnez & S. Altman, (1988) Protein-RNA interactions in the RNase P holoenzyme from *Escherichia coli*. *J. Mol. Biol.* **202**: 835-848.

Vitreschak, A. G., D. A. Rodionov, A. A. Mironov & M. S. Gelfand, (2003) Regulation of the vitamin B_{12} metabolism and transport in bacteria by a conserved RNA structural element. *RNA* **9**: 1084-1097.

Vockenhuber, M. P., C. M. Sharma, M. G. Statt, D. Schmidt, Z. J. Xu, S. Dietrich, H. Liesegang, D. H. Mathews & B. Suess, (2011) Deep sequencing-based identification of small non-coding RNAs in *Streptomyces coelicolor*. *RNA Biol.* **8**: 468-477.

Volker, U. & M. Hecker, (2005) From genomics via proteomics to cellular physiology of the Gram-positive model organism *Bacillus subtilis*. *Cell Microbiol* **7**: 1077-1085.

Vörös, A., B. Horváth, J. Hunyadkürti, A. McDowell, E. Barnard, S. Patrick & I. Nagy, (2012) Complete genome sequences of three *Propionibacterium acnes* isolates from the type IA2 cluster. *J Bacteriol* **194**: 1621-1622.

Vos, M., (2009) Why do bacteria engage in homologous recombination? *Trends in microbiology* **17**: 226-232.

Wada, J., K. Shikata & H. Makino, (1999) Novel approaches to unravel the genesis of glomerulosclerosis by new methodologies in molecular biology and molecular genetics. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association* **14**: 2551-2553.

Walters, C. E., E. Ingham, E. A. Eady, J. H. Cove, J. N. Kearney & W. J. Cunliffe, (1995) *In vitro* modulation of keratinocyte-derived interleukin-1α (IL-1α) and peripheral blood mononuclear cell-derived IL-1β release in response to cutaneous commensal microorganisms. *Infect Immun* **63**: 1223-1228.

Walz, A., V. Pirrotta & K. Ineichen, (1976) Lambda repressor regulates switch between P_R and P_{RM} promoters. *Nature* **262**: 665-669.

Wang, Z., M. Gerstein & M. Snyder, (2009a) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**: 57-63.

Wang, Z., M. Gerstein & M. Snyder, (2009b) RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**: 57-63.

Wanner, U. & T. Egli, (1990) Dynamics of microbial growth and cell composition in batch culture. *FEMS Microbiol Rev* **6**: 19-43.

Wassarman, K. M., (2002) Small RNAs in bacteria: diverse regulators of gene expression in response to environmental changes. *Cell* **109**: 141-144.

Webster, G. F. & E. M. Graber, (2008) Antibiotic treatment for acne vulgaris. *Semin Cutan Med Surg* **27**: 183-187.

Wei, B., Y. Pang, H. Zhu, L. Qu, T. Xiao, H. C. Wei, H. D. Chen & C. D. He, (2010) The epidemiology of adolescent acne in North East China. *J Eur Acad Dermatol Venereol* **24**: 953-957.

Wertz, P. W., M. C. Miethke, S. A. Long, J. S. Strauss & D. T. Downing, (1985) The composition of the ceramides from human stratum corneum and from comedones. *J Invest Dermatol* **84**: 410-412.

Williams, H. C., R. P. Dellavalle & S. Garner, (2012) Acne vulgaris. *The Lancet* **379**: 361-372.

Wright, P. C., J. Noirel, S. Y. Ow & A. Fazeli, (2012) A review of current proteomics technologies with a survey on their widespread use in reproductive biology investigations. *Theriogenology* **77**: 738-765 e752.

Wurtzel, O., R. Sapra, F. Chen, Y. W. Zhu, B. A. Simmons & R. Sorek, (2010) A single-base resolution map of an archaeal transcriptome. *Genome Res.* **20**: 133-141.

Xavier, K. B. & B. L. Bassler, (2003) LuxS quorum sensing: more than just a numbers game. *Curr Opin Microbiol* **6**: 191-197.

Xu, L., H. Li, C. Vuong, V. Vadyvaloo, J. Wang, Y. Yao, M. Otto & Q. Gao, (2006) Role of the *luxS* quorum-sensing system in biofilm formation and virulence of *Staphylococcus epidermidis*. *Infect Immun* **74**: 488-496.

Xue, T., Y. You, D. Hong, H. Sun & B. Sun, (2011) The *Staphylococcus aureus* KdpDE two-component system couples extracellular K^+ sensing and Agr signaling to infection programming. *Infect Immun* **79**: 2154-2167.

Yarchuk, O., N. Jacques, J. Guillerez & M. Dreyfus, (1992) Interdependence of translation, transcription and mRNA degradation in the *lacZ* gene. *J Mol Biol* **226**: 581-596.

Young, R. A. & J. A. Steitz, (1979) Tandem promoters direct *Escherichia coli* rRNA synthesis. *Cell* **17**: 225-234.

Zouboulis, C., A. Eady, M. Philpott, L. Goldsmith, C. Orfanos, W. Cunliffe & R. Rosenfield, (2005) What is the pathogenesis of acne? *Exp Dermatol* **14**: 143-143.

Zouboulis, C. C., (2004) Acne and sebaceous gland function. *Clin Dermatol* **22**: 360-366.

Zouboulis, C. C., (2009) Sebaceous gland receptors. *Dermatoendocrinol* **1**: 77-80.

Zuker, M., (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**: 3406-3415.