

RESEARCH ARTICLE

Open Access

A functional genomics catalogue of activated transcription factors during pathogenesis of pneumococcal disease

Layla K Mahdi^{1†}, Tahereh Deihimi^{2†}, Fatemeh Zamansani^{2†}, Mario Fruzangohar³, David L Adelson⁴, James C Paton¹, Abiodun D Ogunniyi^{1*} and Esmail Ebrahimie⁵

Abstract

Background: *Streptococcus pneumoniae* (the pneumococcus) is the world's foremost microbial pathogen, killing more people each year than HIV, TB or malaria. The capacity to penetrate deeper host tissues contributes substantially to the ability of this organism to cause disease. Here we investigated, for the first time, functional genomics modulation of 3 pneumococcal strains (serotype 2 [D39], serotype 4 [WCH43] and serotype 6A [WCH16]) during transition from the nasopharynx to lungs to blood and to brain of mice at both promoter and domain activation levels.

Results: We found 7 highly activated transcription factors (TFs) [*argR*, *codY*, *hup*, *rpoD*, *rr02*, *scrR* and *smrC*] capable of binding to a large number of up-regulated genes, potentially constituting the regulatory backbone of pneumococcal pathogenesis. Strain D39 showed a distinct profile in employing a large number of TFs during blood infection. Interestingly, the same highly activated TFs used by D39 in blood are also used by WCH16 and WCH43 during brain infection. This indicates that different pneumococcal strains might activate a similar set of TFs and regulatory elements depending on the final site of infection. Hierarchical clustering analysis showed that all the highly activated TFs, except *rpoD*, clustered together with a high level of similarity in all 3 strains, which might suggest redundancy in the regulatory roles of these TFs during infection. Discriminant function analysis of the TFs in various niches highlights differential regulatory backgrounds of the 3 strains, and pathogenesis data confirms *codY* as the most significant predictor discriminating between these strains in various niches, particularly in the blood. Moreover, the predicted TF and domain activation profiles of the 3 strains correspond with their distinct pathogenicity characteristics.

Conclusions: Our findings suggest that the pneumococcus changes the short binding sites in the promoter regions of genes in a niche-specific manner to enhance its ability to disseminate from one host niche to another. This study provides a framework for an improved understanding of the dynamics of pneumococcal pathogenesis, and opens a new avenue into similar investigations in other pathogenic bacteria.

Background

Streptococcus pneumoniae (the pneumococcus) continues to cause high morbidity and mortality worldwide, in spite of the availability of vaccines and antimicrobial therapies [1,2]. Resistance of virulent *S. pneumoniae* to multiple antibiotics, particularly against beta-lactams is due to alterations in the structure of six penicillin-binding

proteins, while macrolide resistance is mediated through the *erm(B)*, *mefA* or *mefE* genes [3]. This could partly be explained by the fact that the organism deploys an efficient virulence machinery during infection. Consequently, research efforts have been geared towards understanding the molecular mechanisms underlying the pathogenesis of pneumococcal disease [4-6]. However, due to technical difficulties associated with harvesting RNA for analysis of *in vivo* transcription patterns of pneumococci during pathogenesis, most studies have either studied the earlier stages of infection (nasopharynx and lungs) or used *in vitro* surrogates [7-10]. Recently, these challenges have been

* Correspondence: david.ogunniyi@adelaide.edu.au

†Equal contributors

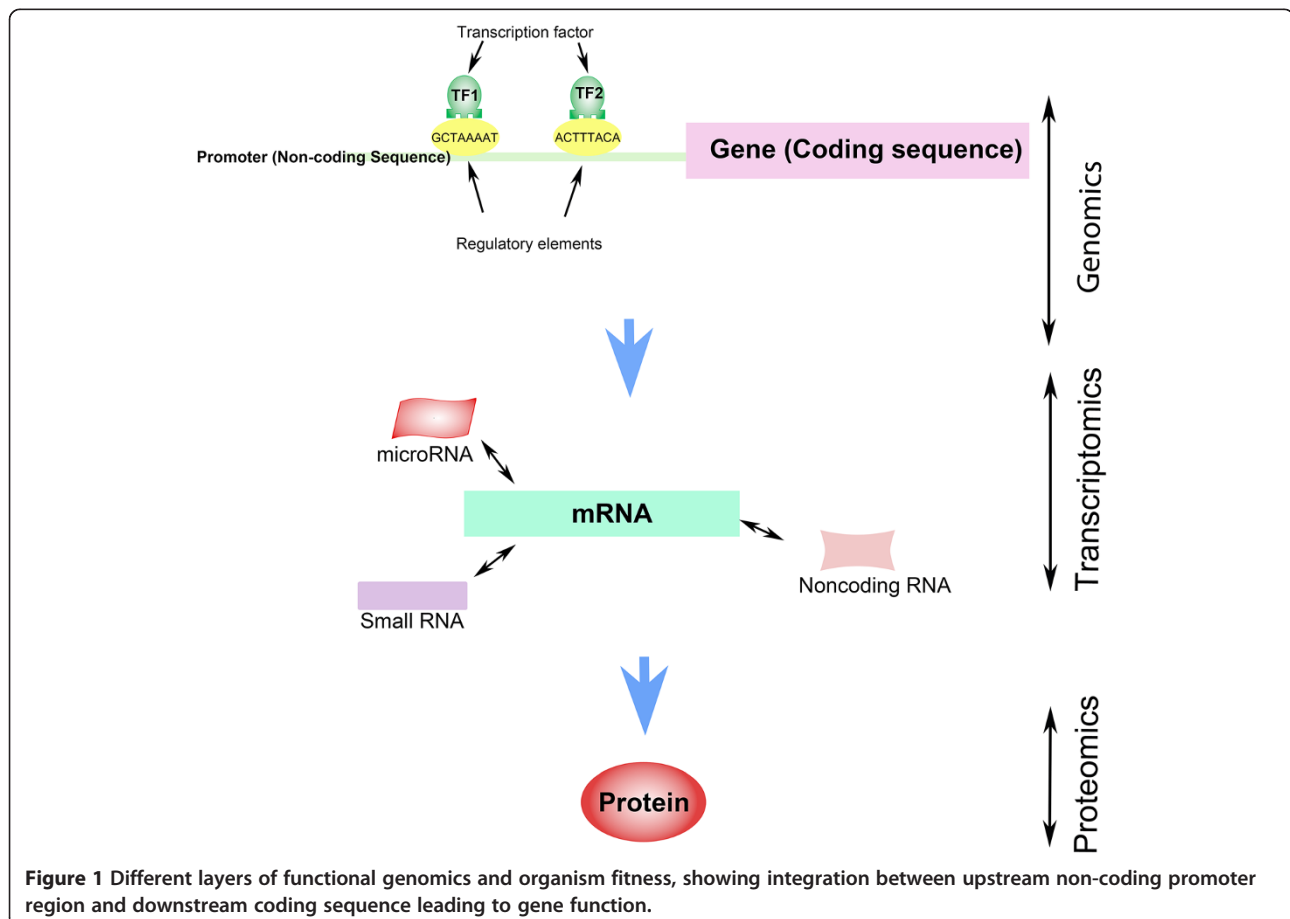
¹Research Centre for Infectious Diseases, School of Molecular and Biomedical Science, The University of Adelaide, Adelaide, South Australia, Australia
Full list of author information is available at the end of the article

largely overcome by studies involving transcriptomic comparisons of differentially-regulated genes during penetration of deeper host tissues [11-13].

Notwithstanding these significant advances, progress on a comprehensive understanding of the dynamics of pneumococcal pathogenesis is still hampered by paucity of data on different levels of functional genomics (such as promoter activation and domain interaction), particularly for pneumococci with distinct pathogenicity characteristics. It is known that gene function is the outcome of harmony between the upstream non-coding promoter region and the downstream coding sequence [14] (Figure 1). Despite the prominent role of transcription factors (TFs) in controlling the expression of many genes [15,16], their impact on pathogenesis of pneumococcal disease has not been studied in detail, probably due to their transient, and generally low expression levels. A study of 60 bacterial genomes showed that larger genomes harbour more TFs per gene than smaller ones [17]. This suggests that under complex conditions, gene expression, regulation and signal integration have been strongly selected to enable rapid adaptation to environmental conditions, triggering emergence of new strains. A differential fluorescence induction

study showed that pneumococcal surface antigen (*psa*) promoter, which drives expression of the *psaBCA* operon involved in manganese uptake and virulence in *S. pneumoniae* [18,19] is markedly activated during lung infection [20]. Recently, we found a novel transcriptional regulatory circuit consisting of two TFs (SmrC [SP_0927] and SP_0676) plays a significant role in pathogenesis and virulence of *S. pneumoniae* [21]. However, to date, there is no report that provides a detailed characterization of TFs (and their corresponding regulatory elements) and their activation patterns during pneumococcal translocation from the nasopharynx to deeper host tissues. The usual approach to address this problem involves the use of *in silico* analysis of promoter regions of differentially regulated genes and prediction of involved TFs, using the whole genomic sequence of a particular strain based on orthology [22-24]. Given the fact that functional specificity of proteins is conserved among orthologs [25], it is possible to compare whole genomes of *S. pneumoniae* with *Escherichia coli* to gain information on the TF and promoter activation map of pneumococcal virulence machinery.

The study of activated domains of expressed proteins in addition to promoter activation profiles can result in



a better understanding of functional genomics. Compared to TFs, more information is available on pneumococcal pathogenesis at the level of expressed protein domains because of its use in vaccine design. The role of ATP-dependent transport and DNA binding domains in pneumococcal competence [26], histidine kinase in sensing environmental cues and virulence [27], and the role of PsaA in virulence [18,28], are well documented. Genome wide analysis of histidine kinases and response regulators has led to the discovery of new response regulator loci [27], while mutation in the DNA gyrase domain has resulted in resistance to quinolones [29].

In the present study, we used functional genomics tools to examine the underlying molecular mechanisms that underpin the different pathogenicity characteristics of 3 pneumococcal strains at both promoter (transcriptomic) and domain (proteomic) activation levels. To our knowledge, this combined approach of characterizing *in vivo* microarray data has not been reported previously and can be used as a model for unraveling functional genomics attributes of other pathogenic bacteria.

Results

For the purposes of this investigation, we have used 3 well-characterised pneumococcal strains: D39 (serotype 2), WCH43 (serotype 4) and WCH16 (serotype 6A). Mouse intranasal challenge experiments have shown that D39 and WCH43 are more virulent than WCH16. However, D39 causes severe pneumonia and high-grade bacteremia, while WCH16 and WCH43 have a propensity to translocate to the brain of infected mice [21,30] (see Additional file 1: Figure S1). Nevertheless, WCH43 infection of mice demonstrates the “classical” disease progression from the nasopharynx to the lungs and dissemination to blood and then to the brain while WCH16 seems to progress directly to the brain with minimal lung and blood involvement [21,31]. These striking differences in the pathogenicity and virulence characteristics of these strains make them ideal for comprehensive functional genomics analyses.

Pneumococcal strains activate a variety of TFs during pathogenesis

Our initial bioinformatic analysis of microarray data for D39, WCH16 and WCH43 at 72 h post-infection of mice suggests that WCH16 up-regulates many genes in the lungs and brain, WCH43 up-regulates many genes in the brain, whereas D39 has significant genome activation in the blood. The list of up-regulated genes of D39, WCH16 and WCH43 during pathogenesis is presented in Additional files 2, 3, 4: Tables S1-S3.

We then characterized pneumococcal TFs capable of binding to a variety of up-regulated genes during pathogenesis and compared their activation profiles between

D39, WCH16 and WCH43. Generally, the highest differential TF activation profile could be observed in the blood (see Additional file 5: Table S4). For D39, a total of 2196 transcription factor-binding sites (TFBs) or regulatory elements were found, while this number dramatically fell to 69 for WCH43 and just 11 for WCH16 in the blood. In this niche, TFs with the highest number (at least 100) of TFBs (such as *argR*, *codY*, *hup*, *rpoD*, *rr02*, *scrR* and *smrC*) were found in D39 (Table 1). The potential TFBs for each of the TFs are listed in Additional file 6: Table S5. The most active TF for all 3 strains in the blood was *rpoD*, although its activation rate is markedly different between the strains (3 predicted binding sites in WCH16, 26 in WCH43, and 767 in D39). Interestingly, a very high number of TFBs were associated with all these TFs (particularly *rpoD*, *hup* and *rr02*) in WCH16 and WCH43 during brain infection. We also compared activated TFs between D39, WCH16 and WCH43 during transition from the nasopharynx to the lungs. This revealed the same 7 activated TFs for WCH16 and WCH43, while 5 were activated for D39. Of these, *rpoD* was the most active TF for all 3 strains, with 74, 23 and 17 binding sites on the promoter regions of up-regulated genes in WCH16, WCH43 and D39, respectively (Table 1).

Activated TFs with the highest number of TFBs cluster together

We sought to determine if there are any relationships or commonality between the activated TFs that we identified above in the 3 strains regarding their binding potential to the promoter regions of up-regulated genes during infection by hierarchical clustering analysis. Interestingly, we found that all activated TFs harbouring the highest number of TFBs (*argR*, *codY*, *hup*, *rr02*, *scrR* and *smrC*), except *rpoD*, clustered together with a high level of similarity (>70%) in all 3 strains (Figure 2, A-D). These results were further validated by cIValid [32], using internal and stability cluster validation measures to design a threshold line to generate 2 clusters. The threshold line shows that *rpoD* has a distinct differential activation profile compared to the other TFs (Figure 2, A-D; see Additional file 7: Table S6).

We also compared the clustering profiles of all activated TFs across niches between the 3 strains. This showed that TFs that regulate the transition of WCH16 and WCH43 from the blood to the brain clustered together, with a very high level of similarity. Moreover, the TFs that regulate translocation from the nasopharynx to the lungs clustered with those activated during transition from the blood to brain for both strains, albeit at a lower level (Figure 3). We also found that TFs that control transition from the nasopharynx to the lungs for D39, and those responsible for transition from the lungs to blood for WCH43 clustered together, while the TFs that regulated

Table 1 Transcription factors (TFs) with the highest number of binding sites on the promoter regions of up-regulated genes during pneumococcal pathogenesis

TF	Lungs vs nasopharynx			Blood vs lungs			Brain vs blood	
	WCH16	WCH43	D39	WCH16	WCH43	D39	WCH16	WCH43
SP_0927 (<i>smrC</i>)	13	4			3	115	18	18
SP_1073 (<i>rpoD</i>)	74	23	17	3	26	767	94	113
SP_1113 (<i>hup</i>)	26	10	4	2	4	249	36	39
SP_1227 (<i>rr02</i>)	20	4	6	1	6	137	31	31
SP_1584 (<i>codY</i>)	19	3			8	175	28	25
SP_1725 (<i>scrR</i>)	16	4	1		1	72	15	17
SP_2077 (<i>argR</i>)	14	7	1	4	7	201	17	23

transition from lungs to blood in D39 were quite distinct. These clustering patterns were also validated by c1Valid, and based on the 2 generated clusters after applying threshold line, blood vs lungs in strain D39 shows a distinct profile of activated TFs (Figure 3, Additional file 7: Table S6). These differences in TF activation profiles across niches are consistent with the bioluminescence patterns of the 3 strains in mice (Additional file 1: Figure S1).

Activated TFs exhibit different behavior between pneumococcal strains

We used discriminant function analysis to compare the relative impact (weight) of TFs between the 3 strains in various niches. Comparison of D39 and WCH16 showed that both strains have relatively similar discriminant models in the lungs. In the blood, a large number of TFs have opposite signs between D39 and WCH16. This highlights differential regulatory backgrounds of both strains

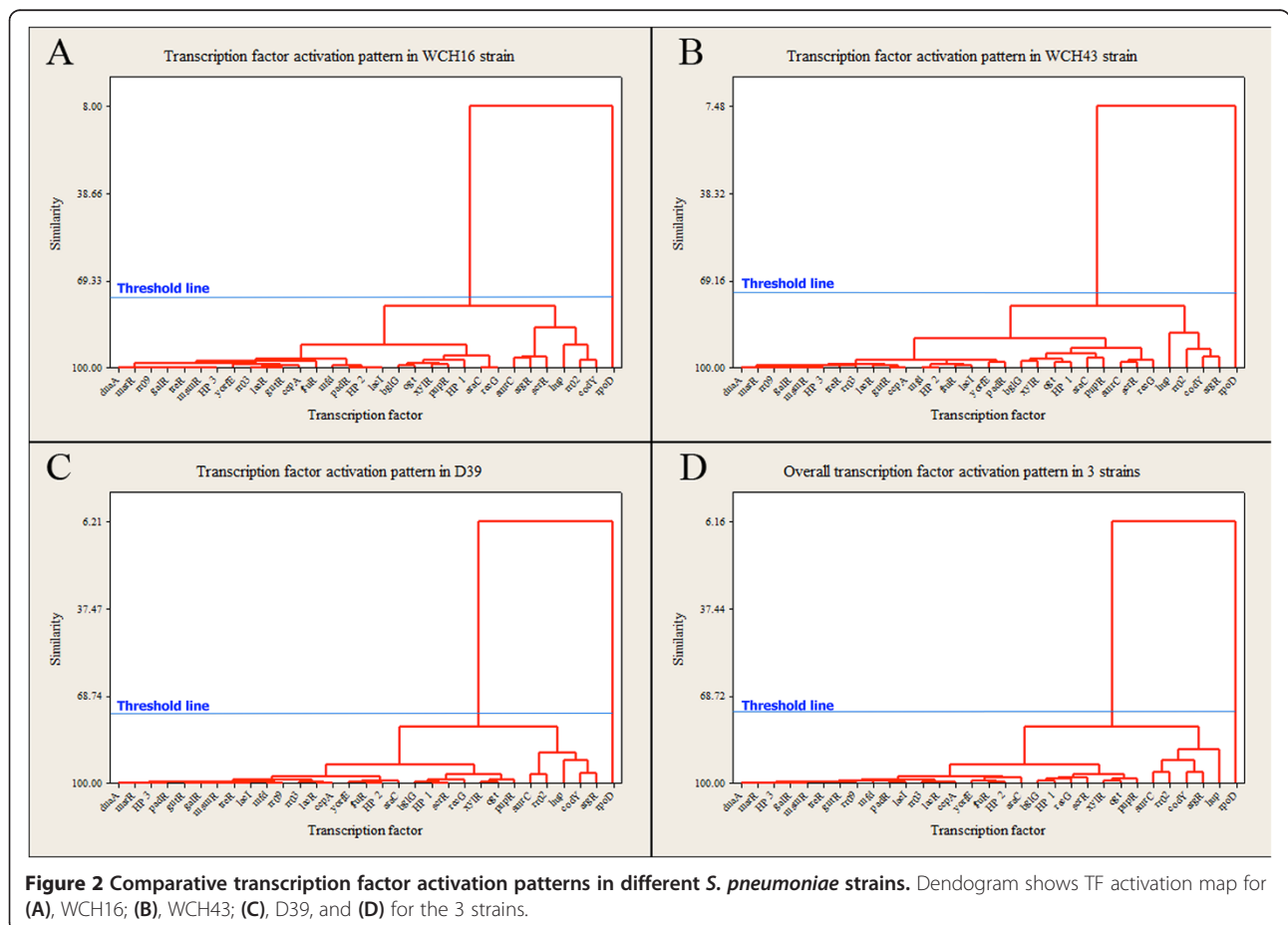
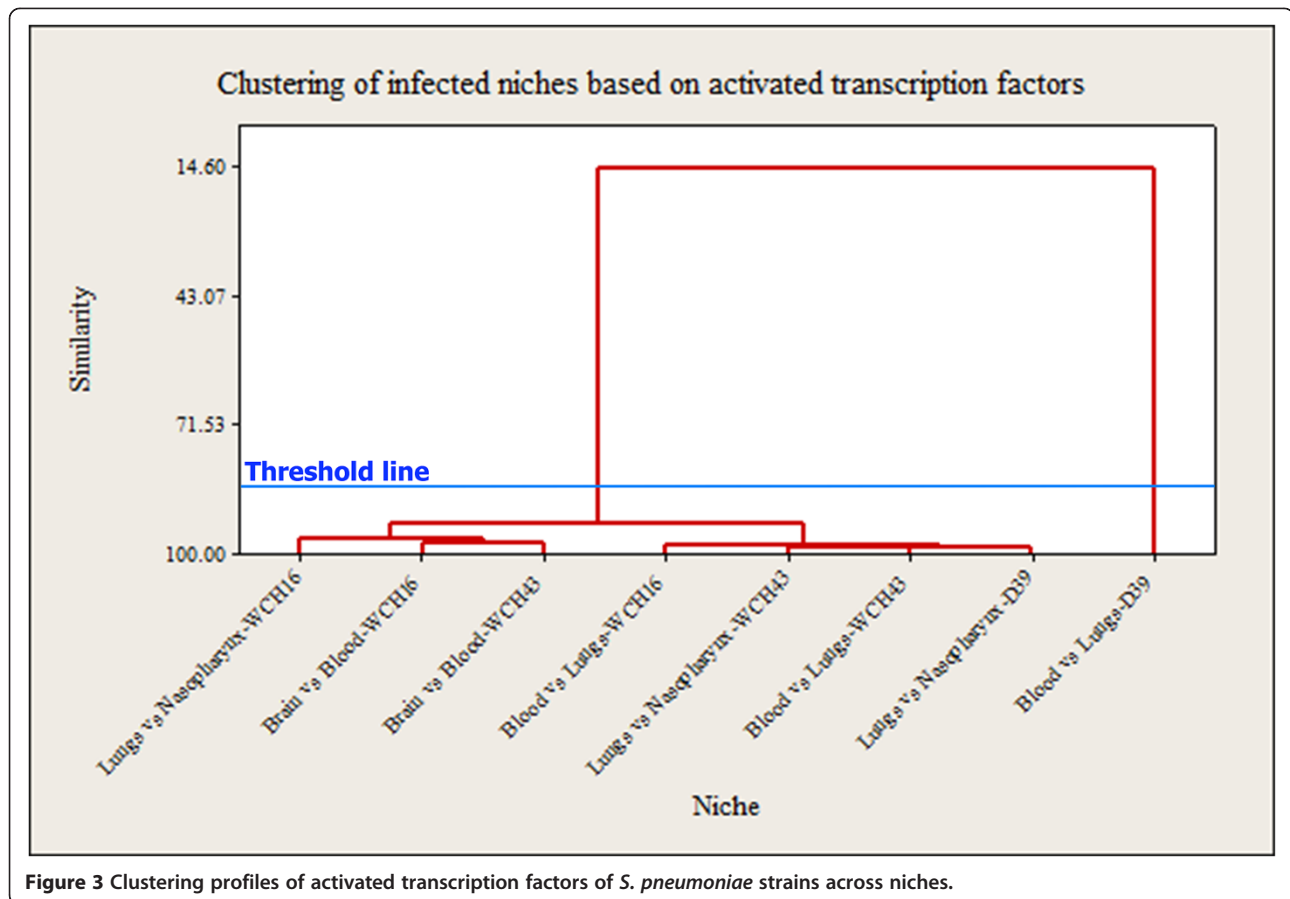


Figure 2 Comparative transcription factor activation patterns in different *S. pneumoniae* strains. Dendrogram shows TF activation map for (A), WCH16; (B), WCH43; (C), D39, and (D) for the 3 strains.



in the blood (see Additional file 8: Table S7). However, the discriminant modelling algorithm indicate that D39 and WCH43 have similar regulatory backgrounds in the lungs but are different in the blood. The discriminant models of WCH16 and WCH43 in the lungs and brain show that the activated TFs have similar coefficient values and signs in both strains (Additional file 8: Table S7). Interestingly, these strains choose different models during progression from lungs to blood. In this instance, of the 7 highly activated TFs, *codY* was the most significant predictor discriminating WCH16 from WCH43 in the blood, where its coefficient was -38 in WCH16, compared to +13 in WCH43.

Given the predicted discriminant function for *codY* between these 3 strains, we hypothesized that a mutant of *codY* will show distinct pathogenic profiles between these strains *in vivo*. However, we noted that a previous manuscript reported that *codY* inactivation in D39 is difficult due to its essentiality [33] and suggested further work in other clinical strains to establish if this is a general feature of pneumococci. Therefore, we attempted to delete *codY* in D39, WCH16 and WCH43 by targeted mutagenesis using the overlap extension strategy, as described in Methods. Surprisingly, we obtained putative mutants in all 3 strains, which were verified to be correct

by PCR and sequencing using primers flanking the *codY* ORF. We examined if the putative *codY* mutants carried the permissive amino acid ($Q_{166} \rightarrow \text{Stop}$) mutation in FatC (SP_1870) and/or amino acids ($D_{480} \rightarrow Y$; $D_{487} \rightarrow Y$) and amino acids 1-81 deletion mutations in AmiC (SP_1890) reported previously [33]. This was carried out by PCR and sequencing of the *amiC* and *fatC* genes of 2 independent mutants from each of the 3 strains. However, we could not detect any of these reported changes. Next, we tested the fitness of each of these putative *codY* mutants in an *in vitro* competition with the isogenic wild type over 4.5 hrs. Surprisingly, we found that the mutant of all 3 strains was completely out-competed by the wild type. Output ratios were 1:1300, 1:655 and 1:1900, for D39 $\Delta codY$, WCH16 $\Delta codY$ and WCH43 $\Delta codY$ mutants, respectively at 3 h. We concluded that while the *codY* mutation might have been tolerated in these strains, the mutation is unstable, in agreement with the finding of Caymaris and colleagues regarding the essentiality of CodY [33]. We also found that during *in vitro* growth, the antibiotic resistance selection marker in the *codY* mutant strains is lost in the absence of selection. This was verified by replica-plating cells grown on antibiotic plates onto plates with or without antibiotic; the

relative rate of loss of selection was more rapid in the WCH16 Δ *codY* (1:6.5), compared to 1:5, and 1:1.4, for the WCH43 Δ *codY* and D39 Δ *codY* putative mutants, respectively.

As a consequence of these findings, we explored an alternative approach to investigating the discriminant function of CodY in the 3 strains. We placed *codY* under the expression of *ami* promoter and transformed this into the 3 strains, generating D39[*codY*]^{ind}, WCH16[*codY*]^{ind}, and D39[*codY*]^{ind}, respectively. Next, we tested the fitness of these *codY*^{ind} strains in an *in vitro* competition experiment with their otherwise isogenic wild-type derivatives over 4.5 hrs. In this model, there was no difference in bacterial counts of the *codY*^{ind} strains compared to their isogenic wild-type counterparts at any of the time points tested (not shown). We then investigated the effect of over-expressing CodY on the *in vivo* fitness of the *codY*^{ind} strains relative to their respective wild type counterparts in a mouse intranasal (i.n.) competition experiment over a 36 h period, as described in Methods. Our analysis shows that there was no significant difference in the ability of the D39[*codY*]^{ind} strain to colonise the nasopharynx or invade the lungs, blood or brain (Figure 4A). However, the WCH16[*codY*]^{ind} strain was massively out-competed by its wild-type counterpart in all these niches (Figure 4B). Interestingly, the WCH43[*codY*]^{ind} strain was only out-competed by the isogenic wild-type strain in its ability to invade the bloodstream and brain (Figure 4C). Together, these analyses confirm the bioinformatics prediction of the discriminant function of CodY in WCH16, consistent with its role in controlling the expression of many genes involved in pneumococcal metabolism and virulence [33,34].

Up-regulated genes are predicted to be co-regulated by similar sets of TFs in various niches

We then attempted to identify genes (with at least 4 TFs in their promoter regions) that are co-regulated by the same 7 highly activated TFs in the lungs, blood or brain. This allowed us to construct a series of TF activation networks for the 3 strains (Figure 5, A-H; see Additional file 9: Table S8). Overall, our results show that the same groups of genes were co-regulated by similar sets of TFs in the lungs and brains of mice infected with WCH16, and in the brains of mice infected with WCH43. These results correlate with the TF activation profiles observed for these strains in those niches (as shown in Figure 3). Interestingly, only SP_1647 (a metallo-endopeptidase; in the lungs of D39-infected mice), and SP_1058 (hypothetical protein; in the blood of WCH16-infected mice) were found to be potentially co-regulated by *argR*, *hup*, *rpoD* and *rr02*. We also observed that D39 showed a distinct set and high number of up-regulated genes in the blood. The majority of these genes are co-regulated by 4 TFs, except

for a set of genes involved in the pentose phosphate pathway (SP_0316-SP_0320) that are co-regulated by all but one (*smrC*) of the 7 highly activated TFs (Figure 5H; Additional file 9: Table S8). A view of the genome maps of sequenced *S. pneumoniae* strains suggests these genes are grouped in transcriptional units, although the sequence of SP_0316 is very short (108 bp) and is absent in most of these strains (except *S. pneumoniae* TIGR4 and *S. pneumoniae* TCH8431/19A). Search for orthologous genes and clustering analysis also shows a high level of identity and conservation for SP_0317-SP_0320 in other streptococci (such as *S. agalactiae*, *S. faecalis*, *S. faecium* and *S. pyogenes*) and in *Bacillus subtilis*. These results are reminiscent of redundancy of TF-TFB interactions as described for overlapping regulons like SoxS, MarA and Rob of *E. coli* [35], and among PrfA/CtsR/HrcA and alternative sigma factors (σ^B , σ^C , σ^H , and σ^I) of *Listeria monocytogenes* [36] and among the ζ^X and ζ^W regulon in *B. subtilis* [37] in which different sites are able to recruit the same TF, while different TFs can recognize similar sites.

Our analysis also included the identification of genes that are commonly up-regulated by each of the 7 highly activated TFs in the lungs, blood or brain for the 3 strains. We restricted our analysis to include the genes that harbor a minimum of 4 TFs in their promoter regions in at least 2 of the 3 strains. This showed that the same set of genes (except SP_0698) are commonly up-regulated in the lungs and brain for WCH16 and WCH43 (see Additional file 10: Table S9). However, these genes were not detected during lung infection by D39. By contrast, only one gene (SP_1329), a putative N-acetylneuraminate lyase, was up-regulated in WCH43 and D39 in the blood. These results are complementary to data presented in Additional file 8: Table S7.

Genes harboring higher numbers of regulatory elements constitute a novel index of quality-based gene selection

We developed a new quality-based selection index based on mining of the number of regulatory elements on the promoter regions of up-regulated genes controlled by the 7 highly activated TFs. Specifically, we hypothesized that genes containing higher numbers of TFs are likely to play key roles in pneumococcal pathogenesis because of their ability to host more TFs. Of the up-regulated genes during infection of lungs and brain by WCH16 and WCH43, SP_0698 (ABC-2 type transport system permease protein) was predicted to contain the highest number of TFs (13) in its promoter region (see Additional file 11: Table S10). In the blood, WCH43 up-regulates SP_2182 (a bacteriocin_IIC-type protein) with 18 TFs in its promoter region. However, in D39, SP_0124 and SP_0125 (bacteriocin_IIC-type proteins), SP_0874 and SP_1036 (hypothetical proteins), and SP_1608 (a DadA-

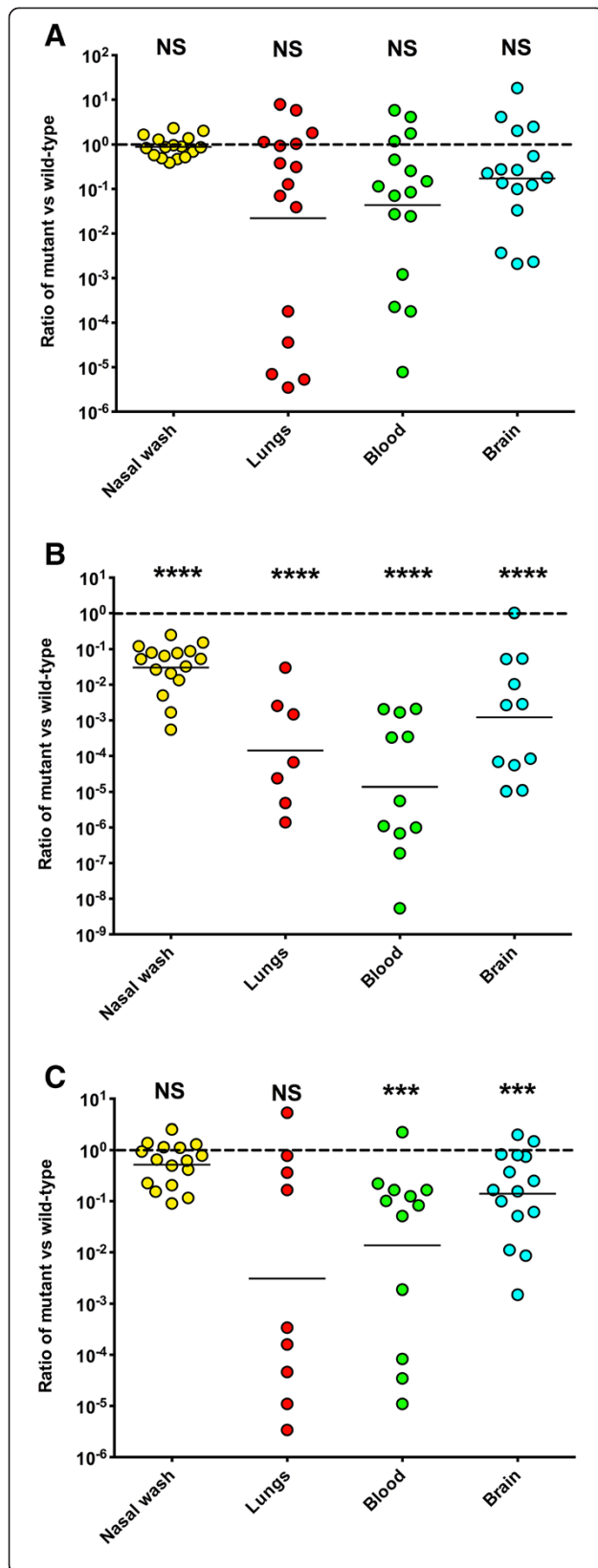


Figure 4 Competition experiments between wild type D39 (A), WCH16 (B), WCH43 (C) and their *codY^{ind}* isogenic counterparts in the nasopharynx, lungs, blood and brain (n = 16) 36 h post-i.n. infection with equal numbers (approx. 2.5×10^6 CFU) each of wild type and isogenic derivative. Datum points are the ratio of *codY^{ind}* bacteria to wild type for each animal. The horizontal broken line represents a 1:1 ratio. The horizontal solid line denotes the geometric mean ratio for each comparison (NS = Not Significant; *** $P < 0.001$; **** $P < 0.0001$; one sample t-test, two-tailed).

family oxidoreductase) (with 23, 26, 25, and 13, TFBs, respectively) were up-regulated.

GO analysis of up-regulated genes containing higher number of TFBs in lungs vs nasopharynx suggests that these genes are involved in key metabolic processes such as NADPH activity, fatty acid biosynthetic process, sugar phosphotransferase system, membrane development, NAD binding, and proteolysis. These processes are important during colonization of the nasopharynx and translocation to lungs by *S. pneumoniae*. In the blood vs lungs comparisons, genes containing high number of TFBs enrich GO processes of ion binding (magnesium ion binding/iron ion binding), transporter activity, metabolic process, as well as ion transmembrane transport. These processes are vital for survival of *S. pneumoniae* in the blood, highlighting the important roles of such up-regulated genes with high TFBs in their promoter regions in the blood. GO analysis of up-regulated genes with high TFBs in the brain vs blood comparison were found to be involved in important processes such as NADPH activity, sugar:hydrogen symporter activity, glutamine biosynthetic process, fatty acid biosynthetic process, sugar phosphotransferase activity, kinase activity, carbohydrate transmembrane transport, ion transmembrane transport, and NAD binding. These processes facilitate the translocation, survival and adaptation of WCH16 and WCH43 in the brain.

Domain activation profiles of different pneumococcal strains during pathogenesis

We reasoned that the study of activated domains of expressed proteins in addition to promoter activation profiles could result in a better understanding of functional genomics because of its use in vaccine design. Therefore, we predicted the complete domain activation catalogue of WCH16, WCH43 and D39 during infection of various host tissues. Accordingly, WCH16 (which is less virulent than WCH43 and D39) is predicted to have the highest amount of its domain activation in the lungs, with 74 predicted domains, compared to 31 for D39 and 23 for WCH43 (Table 2 and Additional file 12: Table S11). Our analysis also suggests that WCH16 has relatively low activation in the blood, with only one predicted activated domain [(N6_N4_Mtase (DNA

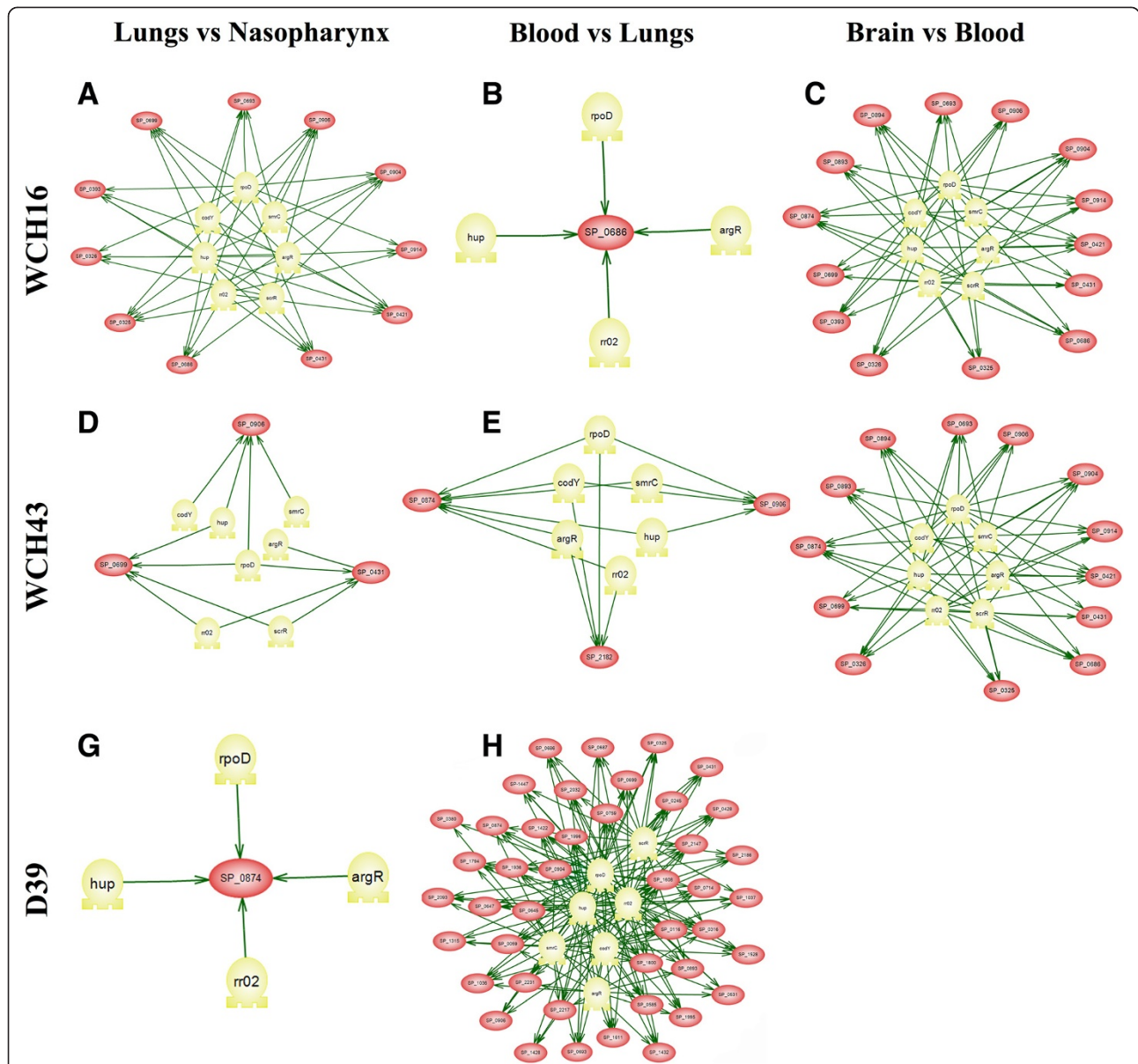


Figure 5 Transcription factor-based regulatory network analysis of *S. pneumoniae* strains during infection. (A) Regulatory network underlying transition of WCH16 from Nasopharynx to Lungs. **(B)** Regulatory network underlying transition of WCH16 from Lungs to Blood. **(C)** Regulatory network underlying transition of WCH16 from Blood to Brain. **(D)** Regulatory network underlying transition of WCH43 from Nasopharynx to Lungs. **(E)** Regulatory network underlying transition of WCH43 from Lungs to Blood. **(F)** Regulatory network underlying transition of WCH43 from Blood to Brain. **(G)** Regulatory network underlying transition of D39 from Nasopharynx to Lungs. **(H)** Regulatory network underlying transition of D39 from Lungs to Blood.

Table 2 Number of activated protein domains by virulence machinery of different *S. pneumoniae* strains during movement from the nasopharynx → lungs → blood → brain

Strain	Lungs vs nasopharynx	Blood vs lungs	Brain vs blood
WCH16	74	1	105
WCH43	23	14	110
D39	31	470	0

methyltransferase)]. In comparison, WCH43 is predicted to activate 14 domains in the blood, while D39 is predicted to exhibit the most remarkable profile, activating 470 possible domains in the blood. It is also predicted that energy producing domains such as ATPase family associated, ATP-synt_ab_N (ATP synthase alpha/beta subunits) and ABC transporters are used by D39 to maintain fulminant bacteremia (Additional file 12: Table S11).

Comparison of the activated domain profiles of WCH16 and WCH43 in the brain revealed that the number of activated domains for both strains are essentially similar (105 and 110, respectively; Table 2). The only slight difference is that in WCH43 more activated domains including energy-producing (such as ATP-grasp), two-component regulatory system, tRNA, ribosomal and PTS systems domains were found.

Discussion

In this work, we have used a functional genomics-based approach to provide a comprehensive prediction of TF activation profiles and domain expression patterns of 3 different *S. pneumoniae* strains during pathogenesis. It has been suggested that the TF and promoter activation map of an organism can be obtained by comparison with the fully annotated map of an ortholog [22-24] given the fact that functional specificity of proteins is conserved among orthologs [25]. Therefore, we exploited the available *E. coli* transcriptional regulation data to gain information on the orthologous TF and promoter activation map of 3 *S. pneumoniae* strains (WCH16, WCH43 and D39) during pathogenesis. However, it has been argued that such methods have shortcomings when applied to distantly related organisms [38]. We attempted to address this potential drawback by using an approach that predicts the function of genes based on the similarity/difference in the pattern of TFs. The identified potential TFs and their organization modules provide another way to understand gene expression and regulation during pathogenesis. Taking into account these approaches, we present a strategy for selection of novel antibacterial targets and vaccine candidates.

In a recent investigation [21], we introduced the concept of bacterial niche-specific virulence gene expression during pathogenesis by qualitative comparisons of transcriptomic data of one niche versus the previous niche. In the present study, we carried out an in-depth analysis of the TF-regulatory network that govern this transition in order to shed light on the virulence factors that *S. pneumoniae* employs to breach host tissue barriers. Our analysis showed that 7 TFs (*argR*, *codY*, *hup*, *rpoD*, *rr02*, *scrR* and *smrC*) were highly activated in all 3 strains *in vivo* and were capable of binding to a large number of up-regulated genes. We suggest that these TFs might constitute the regulatory backbone that underpins pneumococcal pathogenesis. We also found that these TFs were used by D39 in blood and by WCH16 and WCH43 during brain infection.

A close examination of activated TFs of the 3 strains suggests that *S. pneumoniae* adapts to each host niche during infection by increasing the number of TFs deployed (and their TFs). As a corollary, the number of activated TFs is reduced if the bacteria are unable to adapt to a

particular niche (e.g. WCH16 in the blood). Not surprisingly, *rpoD* was the most active TF for all 3 strains in all niches, consistent with its role in promoting the attachment of RNA polymerase to specific initiation sites during transcription. The observed similarity between the predicted TF activation profiles and their clustering patterns between distinct niches suggests that different *S. pneumoniae* strains might activate the same set of TFs and regulatory elements at different sites of infection. This is further supported by discriminant function analysis of the TFs in various niches, which highlights differential regulatory backgrounds of the 3 strains in the blood. We also provide experimental evidence that validates the predicted discriminant function for CodY between the 3 strains by showing distinct pathogenic profiles of their otherwise isogenic *codY* induced derivatives in various anatomical niches of mice. The results are consistent with our *in vivo* microarray and real-time RT-PCR data showing differential expression of CodY-regulated virulence genes such as *aliA*, *ilvH* and *piuA* in various niches of mice [13], and are also in agreement with results from other workers [33]. Furthermore, the predicted domain activation catalogues of these strains during pathogenesis suggests that their functional domain profiles differed mainly in the blood, with only one domain (DNA methyltransferase) activated for WCH16, 14 for WCH43 and 470 domains for D39. These results are consistent with gene ontology classifications for WCH16 and WCH43 [21,39], and in agreement with our published observations showing that these 3 strains display distinct pathogenicity characteristics [21,31,39].

Our analyses also showed that some genes could potentially be regulated by multiple TFs. For example, SP_1647 (a metallo-endopeptidase) was predicted to be regulated by *argR*, *hup*, *rpoD* and *rr02* in the lungs of D39-infected mice. Notably, this gene has recently been characterized to have a role in pneumococcal attachment and internalization in host epithelial (A549) and endothelial (HUVEC) cell lines [40]. We also developed a new quality-based strategy for selection of genes important for pathogenesis, based on mining of the number of TFs in their promoter regions. GO analysis of such genes allowed us to propose that number of TFs in the promoter regions of up-regulated genes is an important and novel index for selection of genes that play key roles at distinct stages of pneumococcal pathogenesis. In this context, we found, amongst other genes, a group of class IIc bacteriocins that harbor the highest number of TFs in WCH43 and D39 harvested from the blood. Bacteriocins have been shown to be involved in fratricide in *S. pneumoniae*, a phenomenon that favors natural competence and pathogenesis in this organism [41,42]. Additionally, a putative N-acetylneuraminase lyase (SP_1329) was up-regulated by both WCH43 and D39 in blood, suggesting a need for this

enzyme for metabolic adaptation of these strains in this niche.

It has been suggested that regulation of pneumococcal virulence proteins is very complex and multifactorial, and likely involves overlapping regulatory mechanisms [43]. We found that all the highly activated TFs (*argR*, *codY*, *hup*, *rrO2*, *scrR* and *smrC*), except *rpoD*, clustered together in all 3 strains. This is a significant finding, implying redundancy in the regulatory roles of these TFs during infection, as the functions of these TFs in pathogenesis and regulation of virulence gene expression have hitherto been studied in isolation [7,33,34,44-47]. Indeed, this notion is exemplified by our recent work, which showed that a defined mutation of *smrC* in a WCH43 background did not completely abrogate virulence [21]. In the same context, we showed in Figure 4 that the same groups of up-regulated genes are predicted to be co-regulated by similar sets of TFs. These findings might explain a fundamental problem associated with prevention and control of pneumococcal disease, as it appears that the task of identifying the principal targets of intervention would be challenging. Therefore, a thorough analysis of the important regulatory pathways employed by *S. pneumoniae* in different *in vivo* niches would likely improve our knowledge of pneumococcal pathogenesis and identify novel targets for intervention.

Conclusions

Our findings suggest a possible evolutionary strategy for the pneumococcus in which the evolution occurs in non-coding promoter regions of genes during infection. Specifically, the pneumococcus changes the short binding sites in the promoter regions of genes instead of alteration of coding sequences in a niche-specific manner to generate more virulent strains. There are two important caveats of this study. Firstly, we utilized the fully annotated *E. coli* promoter listings to pull out the orthologous TFs in *S. pneumoniae*, which might not represent the full array of active promoters in the pneumococcus. Secondly, the microarray slides used in this study cannot detect the expression of unique WCH16-specific genes that are absent in TIGR4 and R6 genomes represented on the slides. Nevertheless, this study is the first such investigation; it provides a framework towards a better understanding of the dynamics of pneumococcal pathogenesis, and opens new avenues into similar investigations in other pathogenic bacteria.

Methods

Bacterial strains and growth conditions

The pneumococcal strains used in this study were D39 (serotype 2; Sequence Type 595), clinical blood isolates WCH43 (serotype 4; Sequence Type 205) and WCH16 (serotype 6A; Sequence Type 4966). Previous mouse

intranasal challenge experiments in our laboratory with these strains indicated that D39 causes fulminant bacteraemia; WCH43 demonstrates the “classical” disease progression from the nasopharynx to the lungs, followed by dissemination to the blood and then to the brain, while WCH16 demonstrates minimal lung and blood involvement before translocation to the brain [31]. Serotype-specific capsule production was confirmed by Quellung reaction, as described previously [48]. Opaque-phase variants of the strains, selected on Todd-Hewitt broth supplemented with 1% yeast extract (THY)-catalase plates [49], were used in all animal experiments. Before infection, the bacteria were grown statically at 37°C in serum broth (SB) to A_{600} of 0.16 (equivalent to approx. 5×10^7 CFU/ml).

Microarray analysis of RNA extracted from different mouse tissues during pathogenesis

For this investigation, we analyzed 72 h-microarray data of *in vivo*-derived RNA samples obtained from our previous studies with WCH16 and WCH43 [12,13] as well as that generated for D39 [12,13,31,50]. Our previous studies have also established that after intranasal infection of CD-1 mice, the number of pneumococci recovered from various tissues peaked at 72 hr after which the mice succumbed to infection quite rapidly [31]. Therefore, we focused on this time point for further differential gene expression analyses. This was carried out by a modified two-color microarray analysis where the relative expression of each gene in one niche was calculated in comparison to expression in the previous niche when bacteria moves from the nasopharynx to lungs to blood and then to brain. Differential gene expression values with $p < 0.05$ were considered as statistically significant (one-sample *t*-test). In the case of multiple comparisons, T- Hotelling using Minitab 16 package (<http://www.minitab.com>) was used to adjust *p*-values and decrease type I error. The goal of this analysis was to unravel pneumococcal gene expression patterns in the nasopharynx, lungs, blood and brain of mice.

Promoter extraction of up-regulated genes

In order to carry out a detailed characterization of the pneumococcal TFs that are active in each host niche during infection, a series of analyses were performed (Figure 6). Firstly, pneumococcal genes that were significantly up-regulated in each niche at 72 h were identified by pairwise comparison of expression levels of each gene in one niche versus the previous niche (e.g. lungs vs nasopharynx, blood vs lungs, brain vs blood). Databases containing full genomic sequences of *S. pneumoniae* TIGR4 and D39, such as Microbesonline (<http://www.microbesonline.org/>) [51,52] and KEGG (<http://www.genome.jp/kegg/>), were used to find the genomic location of up-regulated genes. Then, the potential promoter

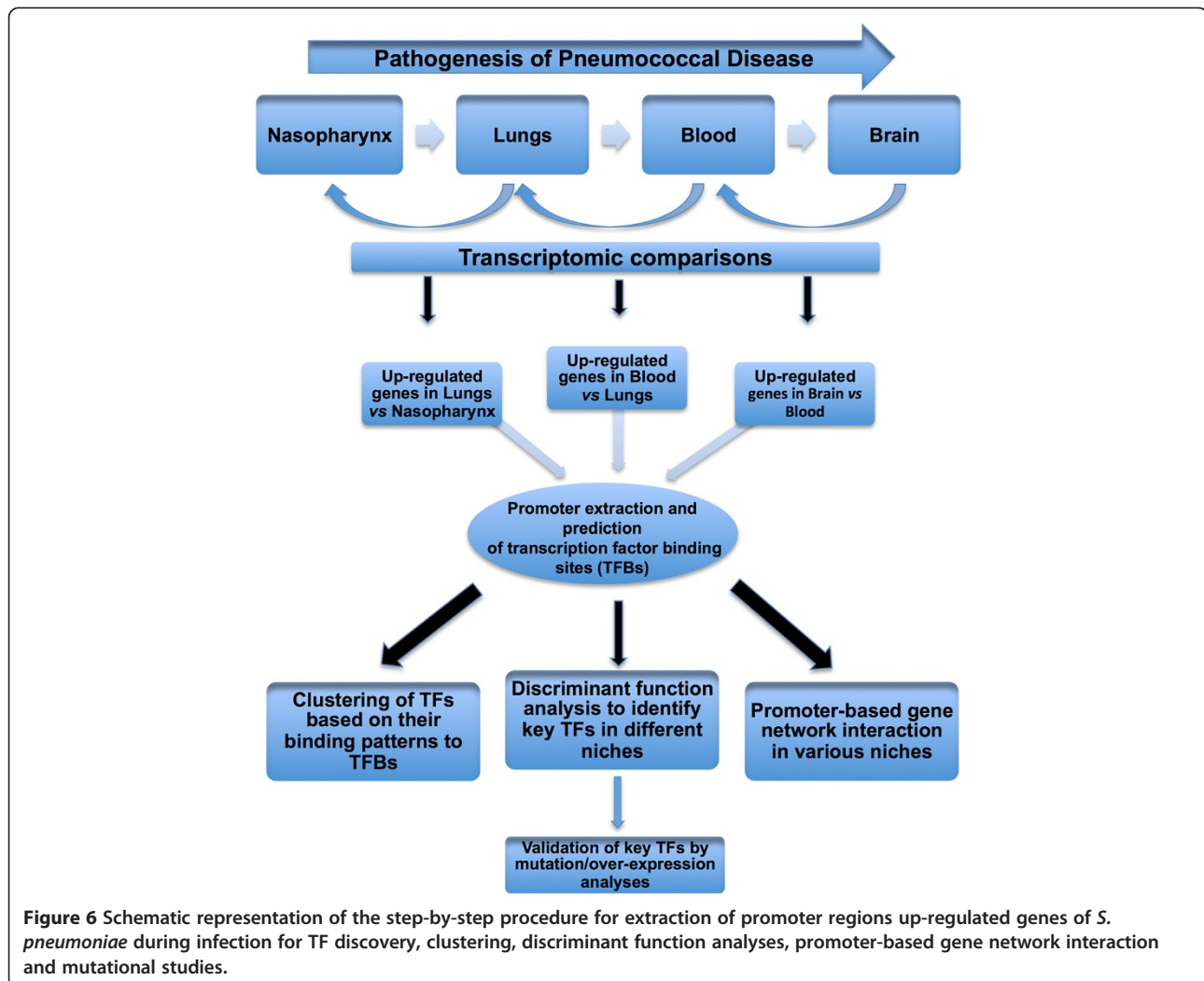


Figure 6 Schematic representation of the step-by-step procedure for extraction of promoter regions up-regulated genes of *S. pneumoniae* during infection for TF discovery, clustering, discriminant function analyses, promoter-based gene network interaction and mutational studies.

regions of up-regulated genes in different niches for all 3 *S. pneumoniae* strains were extracted as follows. The upstream region between each up-regulated gene (or operon) and the next gene (or operon) was extracted from KEGG, Microbesonline database, or Genome-Tools Web Interface (<http://genome-tools.sourceforge.net/cgi-bin/genome-tools-web-interface.pl>) [53]. The promoter analysis covers the entire promoter region of each gene or operon. Existence of a promoter and its -10 and -35 sites was confirmed for up-regulated genes using BPRM algorithm (<http://linux1.softberry.com/berry.phtml?topic=bprom&group=programs&subgroup=gfindb>). The algorithm predicts potential transcription start positions of bacterial genes regulated by sigma70 promoters using linear discriminant function. As the prediction tools for determining TFs are not as complete in *S. pneumoniae*, we used the TFB prediction tool of *E. coli* to predict TFs on the extracted promoter regions for each gene by pattern matching using the BPRM tool at the SoftBerry portal (<http://linux1.softberry.com/berry.phtml>) [54]. TFs were

predicted for all extracted promoter sequences of up-regulated genes in all niches of WCH16, WCH43, and D39 using RegPrecise (<http://regprecise.lbl.gov/RegPrecise/>) [55] and BPRM web applications. The number of regulatory elements (TFBs) in the promoter region of each gene was recorded. In each niche, TFs with the highest number of TFBs in the promoter regions of up-regulated genes were determined for each strain. The number of TFBs for each TF was also recorded. We chose 7 TFs with at least 100 TFBs in their promoter regions as a cut-off for further analysis.

Comparative analysis of TF activation in different niches

Based on promoter analysis of up-regulated genes, a map of TF activation was estimated in different niches for the 3 strains. To obtain a clear understanding of similarities and/or differences in TF activation, TFs were clustered according to the number of TFBs on promoter regions of up-regulated genes during infection by hierarchical clustering analysis (HCA). This was achieved by count

data (number of possible binding sites of each TF to the promoter regions of up-regulated genes in each niche comparison) for all strains. When a particular TF does not have a potential binding site on promoter regions of up-regulated genes, a value of "0" was placed in the data. For each strain, similarities/differences in different TF activation patterns (in terms of potential binding to promoter regions of up-regulated genes) were calculated by average linkage method based on Euclidean distance measure [56]. Dendograms were designed based on calculated similarity matrixes by Minitab 16 package (<http://www.minitab.com>) to visualise differential binding patterns. In addition, clustering of overall TF activation map in all strains and niches was carried out as described above. The hierarchical clustering analyses were validated using c1Valid R Package [32], by internal and stability cluster validation measures. Internal measures represent the compactness, connectedness, and separation of the cluster partitions by Connectivity, Silhouette Width and Dunn Index, where lower Connectivity but higher Dunn and Silhouette represent better scores. Stability measures assess the results from the full clustering data based on removing each column, one at a time. The Stability measures evaluated were: Average proportion of non-overlap (APN), Average distance (AD), Average distance between means (ADM) and Figure of merit (FOM); lower APN, AD, ADM, and FOM represent better scores. We then specified best threshold line to generate 2 clusters.

Discriminant function analysis

We used linear discriminant statistical analysis to determine the functions of TFs and used this information to discriminate between the impact of a particular TF in each niche between two groups (e.g. WCH16 vs WCH43, WCH16 vs D39, WCH43 vs D39). This predictive algorithm provides an accurate comparative index of TF function, whereby TFs with relative higher coefficient values (either positive or negative) have significant influence/impact in a particular niche between the two groups being compared.

GO analysis of up-regulated genes with high number of TFs in their promoter regions

Up-regulated genes containing higher number of TFs during pneumococcal pathogenesis were subjected to GO classification for their functional association categories with respect to molecular function and biological process, using our recently developed bacterial GO web application [39,57].

TF-based network construction of pneumococcal pathogenesis

A database was made based on the results of promoter analysis linking the genes to TFs. TF-based regulatory

networks were designed in different niches for each strain based on the developed database. The database relationships were visualised using PathwayStudio 9 (Elsevier, USA). The database is available upon request.

Extraction of domain profile of up-regulated genes

Protein sequences of up-regulated genes of WCH16, WCH43, and D39 were extracted from PIR (Protein Information Resource) databases (<http://pir.georgetown.edu/cgi-bin/batch.pl>). Then, domains of each protein sequence were extracted from pfam database (<http://pfam.sanger.ac.uk/>) and CLC Genomics Workbench 6 (<http://www.clcbio.com/>). For domain prediction, E-value = 0.0000001 was used to extract the significant domains.

Construction of mutants and assessment of bacterial growth

S. pneumoniae derivatives with marked mutations in *codY* were constructed in the 3 strains. Mutants were constructed by overlap extension PCR [58,59] and validated by PCR and sequencing to be in-frame deletion mutation replacements. All PCR procedures were performed with the Phusion High Fidelity Kit (FINNZYMES). The primer pairs used for construction and validation of the mutants are listed in Additional file 13: Table S12. A recombinant plasmid over-expressing *codY* was constructed by driving its expression from *ami* promoter using pAL2 plasmid [60] in which the *luxABCDE* cassette had been deleted (pAL2 Δ *luxABCDE*). The recombinant construct was transformed into the 3 strains, generating D39[*codY*^{ind}], WCH16[*codY*^{ind}], and D39[*codY*^{ind}], respectively. Induction of *codY* in these strains was confirmed by real-time RT-PCR analysis.

In vitro competition experiments were performed as described previously [21]. In this assay, mutant (or *codY*^{ind}) and wild-type bacteria were grown to A_{600} in SB and then mixed at an input ratio of 1:1 in SB. At 1.5 and 3 h post incubation, an aliquot of each sample was serially diluted in SB and plated on blood agar and blood agar with a selective antibiotic to determine the ratio of mutant (or *codY*^{ind}) to wild-type bacteria. Each competition experiment was repeated at least twice. Competitive indices were calculated as the ratio (\pm SEM) of mutant (or *codY*^{ind}) to wild-type bacteria recovered at each time point adjusted by the input ratio.

Mixed infection experiments

Competition experiments were carried out essentially as described previously [12]. Briefly, 8 mice were anesthetized by intraperitoneal injection of pentobarbital sodium (Nembutal; Rhone-Merieux) at a dose of 66 mg per g of body weight and separately challenged i.n. with 50 μ l suspension containing approx. 2.5×10^6 CFU of either wild-type or the isogenic *codY*^{ind} strain. At 36 h post-challenge, mice from each separate infection experiment were

sacrificed, bacteria were enumerated from the nasopharynx, lungs blood and brain, by plating on blood agar and blood agar with a selective antibiotic to determine the ratio of the *codY*^{ind} strain to wild-type bacteria. Each competition experiment was repeated at least twice. Competitive indices were calculated as the ratio (\pm SEM) of *codY*^{ind} strain to wild-type bacteria recovered in each niche adjusted by the input ratio.

Ethical considerations

Outbred 5- to 6-week-old female CD1 (Swiss) mice were used in all experiments. The Animal Ethics Committee of The University of Adelaide approved all animal experiments (Project Number: S-2013-053). The study was conducted in compliance with the Australian Code of Practice for the Care and Use of Animals for Scientific Purposes (7th Edition 2004) and the South Australian Animal Welfare Act 1985.

Data access

The data reported in this paper are archived at the following databases: BμG@Sbase (<http://bugs.sgul.ac.uk/E-BUGS-130> and <http://bugs.sgul.ac.uk/E-BUGS-133>, and also ArrayExpress (accession number E-BUGS-130, and E-BUGS-133).

Additional files

Additional file 1: Figure S1. Bioluminescent imaging of mice infected with WCH16, WCH43 or D39 at 72 h post-challenge, showing bacteria in the nasopharynx, lungs, blood and brain, for WCH16 and WCH43, and in the nasopharynx, lungs and blood for D39.

Additional file 2: Table S1. List of up-regulated genes of *S. pneumoniae* WCH16 during pathogenesis.

Additional file 3: Table S2. List of up-regulated genes of *S. pneumoniae* WCH43 during pathogenesis.

Additional file 4: Table S3. List of up-regulated genes of *S. pneumoniae* D39 during pathogenesis.

Additional file 5: Table S4. Transcription factor (TF) activation catalogue of 3 pneumococcal strains with different pathogenic profiles during infection.

Additional file 6: Table S5. List of highly activated transcription factors (TFs) during pathogenesis of *S. pneumoniae* WCH16, WCH43, and D39 and their potential regulatory elements.

Additional file 7: Table S6. Validation of TF activation maps for *S. pneumoniae* WCH16, WCH43 and D39, and for all 3 strains and validation of the clustering profiles of activated TFs of all 3 *S. pneumoniae* strains across niches (Figures 2A, 2B, 2C, 2D and Figure 3).

Additional file 8: Table S7. Comparative discriminant function analysis of *Streptococcus pneumoniae* transcription factors during pathogenesis.

Additional file 9: Table S8. Genes co-regulated by the same Transcription factors (TFs) during pneumococcal pathogenesis [SP_0927 (*smrC*), SP_1073 (*rpoD*), SP_1113 (*hup*), SP_1227 (*rr02*), SP_1584 (*codY*), SP_1725 (*scrR*), and SP_2077 (*argR*)].

Additional file 10: Table S9. Transcription factor-specific upregulated genes shared by *S pneumoniae* WCH16, WCH43 and D39 during pathogenesis.

Additional file 11: Table S10. *S. pneumoniae* up-regulated genes under the control of highly activated transcription factors and the number of transcription factor binding sites (TFBs) in their promoter regions.

Additional file 12: Table S11. Domain activation profiles of *S. pneumoniae* WCH16, WCH43, and D39 during transition from the nasopharynx → lungs → blood → brain.

Additional file 13: Table S12. Primers for construction of mutants, cloning, sequencing and real-time RT-PCR analysis.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

LKM, TD and FZ contributed equally to this work. LKM, DLA, JCP, ADO and EE conceived and designed the experiments; LKM, TD, FZ, ADO and EE performed the experiments; LKM, TD, FZ, MF, DLA, JCP, ADO and EE analyzed the data; DLA and EE contributed analysis tools; LKM, MF, JCP, ADO and EE wrote the paper. JCP, ADO and EE are joint senior authors. All authors read and approved the final manuscript.

Authors' information

James C Paton, Abiodun D Ogunniyi and Esmaeil Ebrahimie Joint Senior authors.

Acknowledgements

This work was supported by the Meningitis Research Foundation (UK) Research Grant 0802.0 to ADO, JCP and LKM, the National Health and Medical Research Council of Australia (NHMRC) Project Grant 627142 to JCP and ADO, and NHMRC Program Grant 565526 to JCP. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We thank Kim LeMessurier for help with D39 microarray, and Manije Mohammadi-dehcheshmeh for help with preparation of figures. We also thank Adam Potter for construction of the pAL2Δ*luxABCDE* plasmid. We acknowledge BμG@S (the Bacterial Microarray Guild at St George's, University of London) for supply of the microarray slides and advice and The Wellcome Trust for funding the multi-collaborative microbial pathogen microarray facility under its Functional Genomics Resources Initiative.

Author details

¹Research Centre for Infectious Diseases, School of Molecular and Biomedical Science, The University of Adelaide, Adelaide, South Australia, Australia. ²Institute of Biotechnology, Shiraz University, Shiraz, Iran. ³Australian Centre for Plant Functional Genomics, The University of Adelaide, Waite Campus, Adelaide, South Australia, Australia. ⁴Centre for Bioinformatics and Computational Genetics, School of Molecular and Biomedical Science, The University of Adelaide, Adelaide, South Australia, Australia. ⁵School of Molecular and Biomedical Science, The University of Adelaide, Adelaide, South Australia, Australia.

Received: 25 July 2014 Accepted: 3 September 2014

Published: 8 September 2014

References

1. Paton JC, Boslego JW: **Protein vaccines.** In *Pneumococcal vaccines: The impact of conjugate vaccines.* Edited by Siber GR, Klugman KP, Makela PH. Washington DC: ASM Press; 2008:421–436.
2. O'Brien KL, Wolfson LJ, Watt JP, Henkle E, Deloria-Knoll M, McCall N, Lee E, Mulholland K, Levine OS, Cherian T: **Burden of disease caused by *Streptococcus pneumoniae* in children younger than 5 years: global estimates.** *Lancet* 2009, **374**(9693):893–902.
3. Reinert RR: **The antimicrobial resistance profile of *Streptococcus pneumoniae*.** *Clin Microbiol Infect* 2009, **15**(Suppl 3):7–11.
4. Bogaert D, De Groot R, Hermans PW: ***Streptococcus pneumoniae* colonisation: the key to pneumococcal disease.** *Lancet Infect Dis* 2004, **4**(3):144–154.
5. Kadioglu A, Weiser JN, Paton JC, Andrew PW: **The role of *Streptococcus pneumoniae* virulence factors in host respiratory colonization and disease.** *Nat Rev Microbiol* 2008, **6**(4):288–301.
6. van der Poll T, Opal SM: **Pathogenesis, treatment, and prevention of pneumococcal pneumonia.** *Lancet* 2009, **374**(9700):1543–1556.
7. Hava DL, Camilli A: **Large-scale identification of serotype 4 *Streptococcus pneumoniae* virulence factors.** *Mol Microbiol* 2002, **45**(5):1389–1406.

8. Marra A, Asundi J, Bartilson M, Lawson S, Fang F, Christine J, Wiesner C, Brigham D, Schneider WP, Hromockyj AE: **Differential fluorescence induction analysis of *Streptococcus pneumoniae* identifies genes involved in pathogenesis.** *Infect Immun* 2002, **70**(3):1422–1433.
9. Orihuela CJ, Radin JN, Sublett JE, Gao G, Kaushal D, Tuomanen E: **Microarray analysis of pneumococcal gene expression during invasive disease.** *Infect Immun* 2004, **72**(10):5582–5596.
10. van Opijnen T, Camilli A: **A fine scale phenotype-genotype virulence map of a bacterial pathogen.** *Genome Res* 2012, **22**(12):2541–2551.
11. Molzen TE, Burghout P, Bootsma HJ, Brandt CT, van der Gaast-de Jongh CE, Eleveld MJ, Verbeek MM, Frimodt-Moller N, Ostergaard C, Hermans PW: **Genome-wide identification of *Streptococcus pneumoniae* genes essential for bacterial replication during experimental meningitis.** *Infect Immun* 2011, **79**(1):288–297.
12. Mahdi LK, Wang H, Van der Hoek MB, Paton JC, Ogunniyi AD: **Identification of a novel pneumococcal vaccine antigen preferentially expressed during meningitis in mice.** *J Clin Invest* 2012, **122**(6):2208–2220.
13. Ogunniyi AD, Mahdi LK, Trappetti C, Verhoeven N, Mermans D, Van der Hoek MB, Plumptre CD, Paton JC: **Identification of genes that contribute to the pathogenesis of invasive pneumococcal disease by *in vivo* transcriptomic analysis.** *Infect Immun* 2012, **80**(9):3268–3278.
14. Deihimi T, Niazi A, Ebrahimi M, Kajbaf K, Fanaee S, Bakhtiarzadeh MR, Ebrahimi E: **Finding the undiscovered roles of genes: an approach using mutual ranking of coexpressed genes and promoter architecture-case study: dual roles of thaumatin like proteins in biotic and abiotic stresses.** *SpringerPlus* 2012, **1**:30.
15. Schoolnik GK: **Microarray Analysis of Bacterial Pathogenicity.** In *Advances in Microbial Physiology*, Volume 46. Edited by Poole RK. Elsevier USA: Elsevier; 2002:1–45.
16. Balleza E, Lopez-Bojorquez LN, Martinez-Antonio A, Resendis-Antonio O, Lozada-Chavez I, Balderas-Martinez YI, Encarnacion S, Collado-Vides J: **Regulation by transcription factors in bacteria: beyond description.** *FEMS Microbiol Rev* 2009, **33**(1):133–151.
17. Cases I, de Lorenzo V, Ouzounis CA: **Transcription regulation and environmental adaptation in bacteria.** *Trends Microbiol* 2003, **11**(6):248–253.
18. McAllister LJ, Tseng HJ, Ogunniyi AD, Jennings MP, McEwan AG, Paton JC: **Molecular analysis of the *psa* permease complex of *Streptococcus pneumoniae*.** *Mol Microbiol* 2004, **53**(3):889–901.
19. Ogunniyi AD, Mahdi LK, Jennings MP, McEwan AG, McDevitt CA, Van der Hoek MB, Bagley CJ, Hoffmann P, Gould KA, Paton JC: **Central role of manganese in regulation of stress responses, physiology, and metabolism in *Streptococcus pneumoniae*.** *J Bacteriol* 2010, **192**(17):4489–4497.
20. Marra A, Lawson S, Asundi JS, Brigham D, Hromockyj AE: ***In vivo* characterization of the *psa* genes from *Streptococcus pneumoniae* in multiple models of infection.** *Microbiology* 2002, **148**(Pt 5):1483–1491.
21. Mahdi LK, Ebrahimi E, Adelson DL, Paton JC, Ogunniyi AD: **A transcription factor contributes to pathogenesis and virulence in *Streptococcus pneumoniae*.** *PLoS One* 2013, **8**(8):e70862.
22. Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson BO: **Integrating high-throughput and computational data elucidates bacterial networks.** *Nature* 2004, **429**(6987):92–96.
23. van Schaik W, van der Voort M, Molenaar D, Moezelaar R, de Vos WM, Abee T: **Identification of the σ^B regulon of *Bacillus cereus* and conservation of σ^B -regulated genes in low-GC-content gram-positive bacteria.** *J Bacteriol* 2007, **189**(12):4384–4390.
24. Lintner RE, Mishra PK, Srivastava P, Martinez-Vaz BM, Khodursky AB, Blumenthal RM: **Limited functional conservation of a global regulator among related bacterial genera: Lrp in *Escherichia*, *Proteus* and *Vibrio*.** *BMC Microbiol* 2008, **8**:60.
25. Mirny LA, Gelfand MS: **Using orthologous and paralogous proteins to identify specificity-determining residues in bacterial transcription factors.** *J Mol Biol* 2002, **321**(1):7–20.
26. Hui FM, Morrison DA: **Genetic transformation in *Streptococcus pneumoniae*: nucleotide sequence analysis shows *comA*, a gene required for competence induction, to be a member of the bacterial ATP-dependent transport protein family.** *J Bacteriol* 1991, **173**(1):372–381.
27. Lange R, Wagner C, de Saizieu A, Flint N, Molnos J, Stieger M, Caspers P, Kamber M, Keck W, Amrein KE: **Domain organization and molecular characterization of 13 two-component systems identified by genome sequencing of *Streptococcus pneumoniae*.** *Gene* 1999, **237**(1):223–234.
28. Berry AM, Paton JC: **Sequence heterogeneity of PsaA, a 37-kilodalton putative adhesin essential for virulence of *Streptococcus pneumoniae*.** *Infect Immun* 1996, **64**(12):5255–5262.
29. Pan XS, Ambler J, Mehtar S, Fisher LM: **Involvement of topoisomerase IV and DNA gyrase as ciprofloxacin targets in *Streptococcus pneumoniae*.** *Antimicrob Agents Chemother* 1996, **40**(10):2321–2326.
30. Orihuela CJ, Gao G, McGee M, Yu J, Francis KP, Tuomanen E: **Organ-specific models of *Streptococcus pneumoniae* disease.** *Scand J Infect Dis* 2003, **35**(9):647–652.
31. Mahdi LK, Ogunniyi AD, LeMessurier KS, Paton JC: **Pneumococcal virulence gene expression and host cytokine profiles during pathogenesis of invasive disease.** *Infect Immun* 2008, **76**(2):646–657.
32. Brock G, Pihur V, Datta S, Datta S: **cValid, an R package for cluster validation.** *J Stat Soft* 2008, **25**(4):1–22.
33. Caymaris S, Bootsma HJ, Martin B, Hermans PW, Prudhomme M, Claverys JP: **The global nutritional regulator CodY is an essential protein in the human pathogen *Streptococcus pneumoniae*.** *Mol Microbiol* 2010, **78**(2):344–360.
34. Hendriksen WT, Bootsma HJ, Esteveao S, Hoogenboezem T, de Jong A, de Groot R, Kuipers OP, Hermans PW: **CodY of *Streptococcus pneumoniae*: link between nutritional gene regulation and colonization.** *J Bacteriol* 2008, **190**(2):590–601.
35. Martin RG, Rosner JL: **Genomics of the *marA/soxS/rob* regulon of *Escherichia coli*: identification of directly activated promoters by application of molecular genetics and informatics to microarray data.** *Mol Microbiol* 2002, **44**(6):1611–1624.
36. Chaturongakul S, Raengpradub S, Palmer ME, Bergholz TM, Orsi RH, Hu Y, Ollinger J, Wiedmann M, Boor KJ: **Transcriptomic and phenotypic analyses identify coregulated, overlapping regulons among PrfA, CtsR, HrcA, and the alternative sigma factors σ^B , σ^C , σ^H , and σ^I in *Listeria monocytogenes*.** *Appl Environ Microbiol* 2011, **77**(1):187–200.
37. Huang X, Fredrick KL, Helmann JD: **Promoter recognition by *Bacillus subtilis* sigmaW: autoregulation and partial overlap with the sigmaX regulon.** *J Bacteriol* 1998, **180**(15):3765–3770.
38. Price MN, Dehal PS, Arkin AP: **Orthologous transcription factors in bacteria have different functions and regulate different genes.** *PLoS Comput Biol* 2007, **3**(9):1739–1750.
39. Fruzangohar M, Ebrahimi E, Ogunniyi AD, Mahdi LK, Paton JC, Adelson DL: **Comparative GO: a web application for comparative gene ontology and gene ontology-based gene selection in bacteria.** *PLoS One* 2013, **8**(3):e58759.
40. Agarwal V, Kuchipudi A, Fulde M, Riesbeck K, Bergmann S, Blom AM: ***Streptococcus pneumoniae* Endopeptidase O (PepO) is a multifunctional plasminogen- and fibronectin-binding protein, facilitating evasion of innate immunity and invasion of host cells.** *J Biol Chem* 2013, **288**(10):6849–6863.
41. Guiral S, Mitchell TJ, Martin B, Claverys JP: **Competence-programmed predation of noncompetent cells in the human pathogen *Streptococcus pneumoniae*: genetic requirements.** *Proc Natl Acad Sci U S A* 2005, **102**(24):8710–8715.
42. Weng L, Piotrowski A, Morrison DA: **Exit from competence for genetic transformation in *Streptococcus pneumoniae* is regulated at multiple levels.** *PLoS One* 2013, **8**(5):e64197.
43. Ogunniyi AD, Paton JC: **Regulation of pneumococcal surface proteins and capsule.** In *Regulation of Bacterial Virulence*. Edited by Vasil ML, Darwin AJ. Washington DC: ASM Press; 2012:490–208.
44. Kadioglu A, Echenique J, Manco S, Trombe MC, Andrew PW: **The MicAB two-component signaling system is involved in virulence of *Streptococcus pneumoniae*.** *Infect Immun* 2003, **71**(11):6676–6679.
45. Ng WL, Tsui HC, Winkler ME: **Regulation of the *pspA* virulence factor and essential *pcsB* murein biosynthetic genes by the phosphorylated VicR (YycF) response regulator in *Streptococcus pneumoniae*.** *J Bacteriol* 2005, **187**(21):7444–7459.
46. Iyer R, Camilli A: **Sucrose metabolism contributes to *in vivo* fitness of *Streptococcus pneumoniae*.** *Mol Microbiol* 2007, **66**(1):1–13.
47. Kloosterman TG, Kuipers OP: **Regulation of arginine acquisition and virulence gene expression in the human pathogen *Streptococcus pneumoniae* by transcription regulators ArgR1 and AhrC.** *J Biol Chem* 2011, **286**(52):44594–44605.
48. Berry AM, Paton JC: **Additive attenuation of virulence of *Streptococcus pneumoniae* by mutation of the genes encoding pneumolysin and other putative pneumococcal virulence proteins.** *Infect Immun* 2000, **68**(1):133–140.

49. Weiser JN, Austrian R, Sreenivasan PK, Masure HR: **Phase variation in pneumococcal opacity: relationship between colonial morphology and nasopharyngeal colonization.** *Infect Immun* 1994, **62**(6):2582–2589.
50. LeMessurier KS, Ogunniyi AD, Paton JC: **Differential expression of key pneumococcal virulence genes *in vivo*.** *Microbiology* 2006, **152**(Pt 2):305–311.
51. Dehal PS, Joachimiak MP, Price MN, Bates JT, Baumohl JK, Chivian D, Friedland GD, Huang KH, Keller K, Novichkov PS, Dubchak IL, Alm EJ, Arkin AP: **MicrobesOnline: an integrated portal for comparative and functional genomics.** *Nucleic Acids Res* 2010, **38**(Suppl 1):D396–D400.
52. Alm EJ, Huang KH, Price MN, Koche RP, Keller K, Dubchak IL, Arkin AP: **The MicrobesOnline Web site for comparative genomics.** *Genome Res* 2005, **15**(7):1015–1022.
53. Solovyev V, Salamov A: **Automatic annotation of microbial genomes and metagenomic sequences.** In *Metagenomics and its applications in agriculture, biomedicine and environmental studies*. Edited by Li RW. New York: Nova Science Publishers; 2011:61–78.
54. Lee W, Chen SL: **Genome-Tools: A Flexible Package for Genome Sequence Analysis.** *Biotechniques* 2002, **33**(6):1334–1341.
55. Novichkov PS, Kazakov AE, Ravcheev DA, Leyn SA, Kovaleva GY, Sutormin RA, Kazanov MD, Riehl W, Arkin AP, Dubchak I, Rodionov DA: **RegPrecise 3.0 - A resource for genome-scale exploration of transcriptional regulation in bacteria.** *BMC Genomics* 2013, **14**:745.
56. Manly BFJ: *Multivariate statistical methods: a primer*. 3rd edition. Florida: CRC Press; 2005:214.
57. Fruzangohar M, Ebrahimie E, Adelson DL: **Application of global transcriptome data in gene ontology classification and construction of a gene ontology interaction network.** *bioRxiv* 2014, doi: <http://dx.doi.org/10.1101/004911>.
58. Horton RM, Ho SN, Pullen JK, Hunt HD, Cai Z, Pease LR: **Gene splicing by overlap extension.** *Methods Enzymol* 1993, **217**:270–279.
59. Ogunniyi AD, Grabowicz M, Mahdi LK, Cook J, Gordon DL, Sadlon TA, Paton JC: **Pneumococcal histidine triad proteins are regulated by the Zn²⁺-dependent repressor AdcR and inhibit complement deposition through the recruitment of complement factor H.** *FASEB J* 2009, **23**(3):731–738.
60. Beard SJ, Salisbury V, Lewis RJ, Sharpe JA, MacGowan AP: **Expression of lux genes in a clinical isolate of *Streptococcus pneumoniae*: using bioluminescence to monitor gemifloxacin activity.** *Antimicrob Agents Chemother* 2002, **46**(2):538–542.

doi:10.1186/1471-2164-15-769

Cite this article as: Mahdi et al.: A functional genomics catalogue of activated transcription factors during pathogenesis of pneumococcal disease. *BMC Genomics* 2014 **15**:769.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

