Moonlighting proteins activate transformation in multidrug-resistant *Streptococcus pneumoniae* epigenetic phase variants

Min Jung Kwun[1], Alexandru V. Ion[1], Marco R. Oggioni[2,3], Stephen D. Bentley[4], Nicholas J. Croucher[1,*]

[1] MRC Centre for Global Infectious Disease Analysis, Sir Michael Uren Hub, White City Campus, Imperial College London, London, W12 0BZ, UK

[2] Department of Genetics, University of Leicester, University Road, Leicester, LE1 7RH

[3] Dipartimento di Farmacia e Biotecnologie, Università di Bologna, Via Irnerio 42, 40126 Bologna, Italy

[4] Parasites & Microbes, Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SA, UK

Abstract

Transformation of *Streptococcus pneumoniae* GPSC1 has enabled vaccine evasion and the acquisition of antibiotic resistance. Epigenetic phase variants of GPSC1 isolate RMV7, resulting from rearrangements at the *tvr* restriction-modification locus, differed ~100-fold in their transformation efficiency. This variation was recapitulated in knock-in mutants of the relevant *tvr* alleles. RNA-seq showed the difference was due to blocking of the early competence regulatory cascade. The more transformable variant more highly expressed *manLMN*, encoding a carbon source importer. This was shown to be necessary for efficient competence induction, despite being dispensable for growth in rich media. Transformation was promoted by import of *N*-acetylglucosamine, which activated competence through an orthologue of the gram-negative competence regulator TfoX, and an enzyme likely involved in nucleotide-mediated signalling. The less transformable variant more highly expressed mobile element genes, causing a protein misfolding stress response mediated by HrcA. This chaperone regulator caused the decrease in transformation efficiency associated with heat shock, and the increase associated with elevated Ca^{2+} concentrations. Hence both ManLMN and HrcA moonlighted as activators of competence. Such proteins, conserved by selection on their primary function, are unlikely to be altered in sequence by transformation, and therefore may gain the greatest net benefit from promoting competence.

Competence for natural transformation was first identified in *Streptococcus pneumoniae* (the pneumococcus) in the early 20th century [1]. Cells can be "transformed" to express a new phenotype through the acquisition of exogenous DNA, integrated into their genome through homologous recombination following its import from the environment through the specialised cell-encoded competence machinery [2]. Transformation has played a key role in the emergence of antibiotic-resistant *S. pneumoniae*, both through generating "mosaic" alleles of core loci [3–5] and the acquisition of specialised resistance genes [6]. It has also enabled vaccine evasion through recombinations affecting the capsule polysaccharide synthesis (*cps*) locus altering surface antigens [7,8].

Despite the ability of transformation to accelerate such adaptive evolution in *S. pneumoniae*, considerable variation in the rate of diversification of strains through this mechanism persists across the species [9]. Epidemiological studies have found the *r*/*m* ratio of base substitutions introduced through homologous recombination, relative to those occurring through point mutation, varies from well over 10 [7,9] to below 0.1 [9,10] across the species. Similarly, *in vitro* assays have identified >100-fold differences in the transformation efficiency of *S. pneumoniae* genotypes, with substantial variation even between isolates of the same serotype or strain [11–14]. Many isolates are routinely found not to be transformable under standard conditions [11]. This is often the consequence of integrative mobile genetic elements (MGEs) reducing transformation efficiencies [6,15–17], likely to prevent themselves being eliminated from the chromosome [17], yet in other non-transformable isolates the highly-conserved competence machinery is intact [11,18]. This suggests the variation in transformation rates also reflects differences in activation of the competence system.

The best-characterised stimulus for transformation in *S. pneumoniae* is the competence stimulating peptide (CSP) pheromone, which acts as a quorum-sensing system that signals through the ComDE two-component regulator [19]. This activates early competence genes after about ten minutes [20]. These include *comX*, encoding an alternative sigma factor [21]. ComX enables the RNA polymerase to recognise late competence genes [20], which feature a "combox" signal in their promoters [22,23]. This results in pneumococci entering a transient competent state around 20 minutes post-CSP induction, after which the relevant machinery is degraded [24], and the cells become temporarily refractory to induction [21].

Transformation efficiency is known to vary between isogenic pneumococci through phase variation in capsule production. Transparent colony variants produce less capsule, are less virulent, but more transformable, than opaque colony variants [25]. This short-term variation

has been linked to rapid changes at the phase-variable inverting variable restriction (*ivr*) locus, encoding the conserved Type I *Spn*III restriction-modification system (RMS) and the IvrR recombinase that drives sequence inversions within the locus [26–31]. These rearrangements alter the target recognition domains (TRDs) within the active HsdS specificity protein, which determines the DNA motif that is targeted by both the methylase and endonuclease activities of the system. Consequently, changes at this single locus affect genome-wide methylation patterns, causing pleiotropy through its impact on the transcription of multiple genes [27]. These phase-variable RMSs can thereby maintain phenotypic heterogeneity within a genetically near-homogenous population [32], resulting in “bet hedging” that can increase the chances of a species surviving a changing environment [33].

The second pneumococcal phase-variable Type I restriction-modification system (named *Spn*IV), encoded by the translocating variable restriction (*tvr*) locus, varies through TvrR recombinase-mediated excision-reintegration [27,28,34,35]. This locus is generally conserved across the species, but the complement of TRDs varies between strains, increasing the range of HsdS proteins, and possible methylation patterns, across the species [28,34]. This locus is inactive in the R6x laboratory strain that is typically used to study pneumococcal competence [28,34,36]. Here, we test pairs of otherwise-isogenic *tvr* variants to identify the regulatory mechanisms underlying phenotypic heterogeneity in genotypes commonly causing pneumococcal disease.

Results

Phase variants of *S. pneumoniae* RMV7 differ in their transformation efficiency

Four pairs of otherwise-isogenic “locked” phase variants, in which *tvrR* was disrupted by a selectable and counter-selectable Janus cassette marker [34,37], were screened for differences in their transformation efficiency (Fig. 1A). In the RMV6 and RMV7 pairs, the variant in which the active *tvr* locus HsdS comprised the TRDs III-i (directing the SpnIV system to target the bipartite motif TGAN$_7$TATC) was found to have significantly higher transformation efficiency following induction by exogenous CSP. These closely-related genotypes both originate from the multidrug-resistant GPSC1 strain [38]. Many replicates were required to robustly establish the difference between the RMV6 variants, whereas the RMV7 variants exhibited a clearer ~100-fold difference in their transformation rate (Fig. 1A). This was because the RMV7 variant carrying the alternative form of the *tvr* locus, with an active HsdS comprising the TRDs III-iii (which directs the SpnIV system to target the motif TGAN$_7$TCC; Fig. 1B and S1) had an almost-undetectable transformation efficiency. Culturing the wild-type isolate (RMV7$_{wt}$) over successive days demonstrated large changes in the

relative frequency of these variants, although the less-transformable variant (RMV7 *tvrR*::Janus; henceforth, $RMV7_{domi}$, carrying tvr_{domi}) was typically more common than more transformable variant (RMV7 Δ*tvrR*; henceforth, $RMV7_{rare}$, carrying tvr_{rare}; Fig. 1C). A comparison of $RMV7_{wt}$ and $RMV7_{domi}$ found them to have similarly low transformation efficiencies, whereas that of $RMV7_{rare}$ was confirmed to be ~50-fold higher, consistent with the relative proportions of the variants observed *in vitro* (Fig. 1D). Therefore, $RMV7_{domi}$ and $RMV7_{rare}$ exhibited distinctive phenotypes that correlated with their *tvr* locus arrangements.

$RMV7_{rare}$ was also significantly more adhesive to an abiotic surface (Fig. 1E), which can be considered a proxy for biofilm formation [39]. This could be consistent with $RMV7_{rare}$ being enriched for transparent phase variants. However, microscopy found no clear difference in colony morphology between the variants (Fig. S2). An alternative explanation for the phenotypic differences could be mutations that occurred during genetic manipulation of the isolates [40]. Alignment of the two variants' assemblies found them to be distinguished by seven non-synonymous SNPs and two premature stop codons outside the *tvr* locus, none of which were within genes known to directly affect the competence machinery (Table S1). Nevertheless, we tested the effect on transformation of mutations in $RMV7_{rare}$ that were absent from both the $RMV7_{domi}$ and $RMV7_{wt}$ sequences. A non-synonymous change in *phoB*, encoding a phosphate-sensitive response regulator [41], was not found to affect transformation efficiency (Fig. S3). Similarly, a premature stop codon in *pstS* was observed in $RMV7_{rare}$, but knocking out this gene also failed to explain any difference in transformability between the variants (Fig. S3). Hence the differences between RMV7 variants could not be explained by point mutations or alterations in encapsulation.

To test whether the phenotypic differences were causatively associated with variation in the SpnIV RMS, the *tvr* loci of $RMV7_{domi}$ and $RMV7_{rare}$ were introduced into $RMV7_{wt}$ *tvr*::*cat*. This generated the otherwise isogenic knock-in recombinants RMV7 tvr_{domi}::Janus and RMV7 tvr_{rare}::Janus, carrying two different locked *tvr* loci (Fig. S4). The transformability of both was assayed over five replicate six-day passages. This demonstrated a persistent distinction in the transformation efficiency of the two *tvr* genotypes, which replicated the difference between the original variants (Fig. 1F). Therefore, the RMV7 tvr_{rare} and tvr_{domi} alleles were associated with significant differences in transformation efficiency in otherwise isogenic isolates.

Induction of early and late competence genes is blocked in $RMV7_{domi}$

To understand how the *tvr* loci caused a difference in transformation, RNA-seq was used to quantify patterns of transcription in the recombinants RMV7 tvr_{domi}::Janus and RMV7

tvr$_{rare}$::Janus. Samples were taken pre-CSP, 10 minutes post-CSP, and 20 minutes post-CSP from each of three biological replicates (Fig. S4). The 18 RNA samples were sequenced as 200 nt paired-end multiplexed libraries on a single Illumina HiSeq 4000 lane. Curation of the RNA-seq data found the fragment length distributions (Fig. S5) and gene expression densities (Fig. S6) were consistent across samples (Table S2). Q-Q plots suggested a Benjamini-Hochberg corrected *q* value of 10^{-3} was an appropriate threshold for identifying significant transcriptional variation (Fig. S7-8). This identified 154 genes that significantly differed in their expression between the two genotypes prior to CSP exposure, or between the pre- and post-CSP samples from the same genotype (Fig. 2 and S9-10; Table S3).

Clustering of the replicates found the biggest separation distinguished the post-CSP RMV7 *tvr*$_{rare}$::Janus transcriptomes from the others (Fig. S11), suggesting a major difference in the induction of the competence system between the variants. In RMV7 *tvr*$_{rare}$::Janus, the early competence genes showed elevated expression 10 minutes post-CSP (Fig. S12), with the late competence genes exhibiting more variable patterns of transcription (Fig. S13). Also upregulated was *briC*, a competence-induced biofilm formation signal [42], and some nucleotide metabolism and transporter genes (*purA*, *tadA*, *dut*, *ribF*, *adeQ*; Fig. S14). Another upregulated transporter, encoded by a three gene cluster, had previously been induced in response to exposure to CSP [43] and antimicrobial peptides [44], and consequently was named *pieABC* (peptide-induced exporter; CDSs IONPJBJN_01324-6 in RMV7$_{domi}$, corresponding to SP_0785-787 in TIGR4; Fig. S14, Table S3).

By contrast, CSP upregulated just five genes in RMV7 *tvr*$_{domi}$::Janus: the quorum sensing genes *comCDE* and *briC* (Fig. S12), and a gene encoding a CsbD stress response protein that was not CSP-responsive in RMV7 *tvr*$_{rare}$::Janus. However, the induction of *comCDE* was much weaker, and there was no sign of late competence genes being activated by ComX. Expression of the *comX* gene itself can be difficult to determine through RNA-seq, owing to the presence of two near-identical paralogues in pneumococcal genomes [45].

As an independent test of these transcriptional differences, qRT-PCR experiments were undertaken on the original RMV7$_{domi}$ and RMV7$_{rare}$ variants, and the control genotypes RMV7$_{wt}$ and RMV7$_{wt}$ *tvr*::*cat*, the latter of which lacked a *tvr* locus. Transformation assays performed in parallel with these RNA extractions found RMV7$_{wt}$ *tvr*::*cat* was more efficiently transformable than RMV7$_{wt}$ or RMV7 *tvr*$_{domi}$::Janus, demonstrating the active, dominant form of the locus was repressing transformation (Fig. S4). The qRT-PCR data showed genotypes of both low (RMV7$_{wt}$ and RMV7$_{domi}$) and high (RMV7$_{rare}$ Δ*tvrR* and RMV7$_{wt}$ *tvr*::Janus)

transformation efficiency up-regulated the early competence genes *comD* and *comX* in response to CSP (Fig. S15). However, the late competence genes *comEA* and *comYC* were only significantly upregulated in the highly transformable strains. Hence the difference in transformability between the RMV7 variants was a consequence of late competence gene activation being blocked in RMV7$_{domi}$ through consequences of *tvr*$_{domi}$ expression.

Differences in pre-CSP gene expression

The two pre-CSP samples were compared to identify the cause of this difference in competence induction. Pre-CSP, 33 genes exhibited higher expression in RMV7 *tvr*$_{domi}$::Janus, and 20 genes exhibited higher expression in RMV7 *tvr*$_{rare}$::Janus (Fig. 2; Table S3). These did not include any *cps* locus genes, which were slightly more highly expressed in RMV7 *tvr*$_{rare}$::Janus (Fig. S16), confirming that the elevated transformation efficiency of this genotype did not reflect an enrichment of transparent phase variants [27,46]. To test whether these differentially-expressed genes were affected by methylation within regulatory or promoter sites, the distances from protein coding sequence start codons to the nearest methylation site was plotted for both SpnIV motifs (Fig. S17). There was no general relationship between differential expression and proximal methylation, although four cases of variable methylation sites being within 100 bp of a differentially-expressed gene's start codon were identified. In three cases, the coding sequences close to *tvr*$_{domi}$ motifs were within a larger set of co-transcribed genes, and it was unlikely that transcription initiation would be affected by methylation (Fig. S18). However, one *tvr*$_{rare}$ motif was within the promoter of the *piuABCD* operon (Fig. S18), encoding an iron transporter, that was upregulated in RMV7 *tvr*$_{rare}$::Janus (Fig. S14). However, knocking out *piuA* did not reduce the transformation efficiency of RMV7$_{rare}$, suggesting this change was independent of those affecting competence (Fig. S15).

As the induction of transformation has been suggested to represent a stress response [47], the overall distribution of the alternative SpnIV target motifs was analysed, to test whether differing spatial arrangements across the chromosome might affect cell physiology. While the *tvr*$_{rare}$ motifs were uniformly distributed, the RMV7$_{domi}$ motifs were enriched in one segment of the chromosome (Fig. S20). This suggested that the differences between the two variants was not the result of strong effects at a small number of promoters, but instead a broader response to a general change in the distribution of DNA modifications, as suggested for other epigenetic phase variants [27].

Consistent with the methylation pattern of RMV7$_{domi}$ reducing transformation efficiency through altering cell physiology, many of the differentially-expressed loci identified by RNA-

seq encoded regulators of cell physiology. The *mgrA* gene, encoding a regulator of cell adhesion, was expressed more highly in $RMV7_{domi}$ (Fig. S21). Consistent with previous observations, this correlated with repression of the *rlrA* pilus islet [48] and activation of the *accBC-yqhY-amaP-csbD* cell wall stress operon (IONPJBJN_01032-6) [49] in $RMV7_{domi}$. The cell-surface chaperone-encoding gene *prsA* was also upregulated, as was the chaperone regulator gene *hrcA*. The *groES-groEL* and *grpE-dnaK*-IONPJBJN_02152-*dnaJ* operons of the *hrcA* regulon were also more highly expression in $RMV7_{domi}$, although this only achieved genome-wide significance for IONPJBJN_02152 (Fig. S22; Table S3).

Another indicator of cell stress in $RMV7_{domi}$ was the increased activity of a phage-related chromosomal island (PRCI; also known as a phage-inducible chromosomal island, or PICI) integrated adjacent to *dnaN* ($PRCI_{dnaN}$; Fig. S23; Table S3). This is one of two PRCIs associated with GPSC1 [7,28], the other being integrated near *uvrA* (Fig. 2). The regulatory mechanisms of these elements are not thoroughly characterised [50], and in the absence of a helper prophage, it is unclear exactly what signal may have triggered this increased activity. Given integrative MGEs are likely under selection to reduce host cell transformability [17], the increased activity of $PRCI_{dnaN}$ could drive inhibition of the competence machinery. A further indicator of cell stress was the increased expression of the *ciaRH* two-component system, which is known to inhibit competence [51]. Hence $PRCI_{dnaN}$ and *ciaRH* were the primary candidates for causing the observed difference in transformability between the RMV7 variants.

ManLMN links competence induction to sugar metabolism

The higher expression of the *ciaRH* genes in the non-transformable genotypes $RMV7_{domi}$ and $RMV7_{wt}$, relative to the transformable genotype $RMV7_{rare}$, was confirmed by qRT-PCR (Fig S15). CiaRH binds at least 15 promoter sequences, five of which drive the expression of *cia*-dependent small RNAs (csRNA) that suppress the induction of competence for transformation through inhibiting expression of the *comC* gene, encoding CSP [52]. The remaining ten drive the expression of protein coding sequences (CDSs) [52]. Although csRNA5 was the only non-coding sequence that appeared to respond to *ciaRH* activation, many of the CiaRH regulon CDSs were more highly expressed in $RMV7_{domi}$ compared to $RMV7_{rare}$ pre-CSP (Fig. S24). These included the extracytoplasmic chaperone and serine protease HtrA [52–54], which can block competence induction at low CSP concentrations through degrading extracellular CSP [53,55]. In agreement with some previous studies, elimination of *htrA* further increased the transformability of $RMV7_{rare}$ [53], but the same mutation had no significant effect in $RMV7_{wt}$ (Fig S25). This suggests HtrA inhibits induction of competence, but is unlikely to explain much of the difference between these variants.

Similarly, knock out of *ciaRH* itself had little effect in either genotype (Fig. S25), suggesting an alternative locus was causing the difference in transformation efficiency between the variants.

Among the CiaRH regulon, the most significant difference in expression between $RMV7_{rare}$ and $RMV7_{domi}$ was the ManLMN carbon source importer operon (Table S3). The *manLMN* genes were more highly expressed in $RMV7_{rare}$, as CiaRH binds promoter sequences within the operon [36,52,56,57] and acts as a repressor [58]. Disrupting the *manLMN* locus reduced the transformation efficiency of $RMV7_{rare}$ by >5-fold, while having little effect in $RMV7_{wt}$ (Fig. 3A-B). To test whether the observed transformation differences reflected a growth defect, *manLMN*::Janus mutants of both $RMV7_{wt}$ and $RMV7_{rare}$ were cultured in the same rich media (Fig. S26). However, removal of *manLMN* had little effect on growth, suggesting the transporter's effect on transformation was through regulation rather than proliferation. Hence the lower expression of *manLMN* in $RMV7_{domi}$ accounts for some of its reduced transformation efficiency relative to $RMV7_{rare}$.

ManLMN is a phosphotransferase system (PTS) transporter that can facilitate the import of glucose, mannose, galactose, fructose, aminoglucose (GlnN) and *N*-acetylglucosamine (GlcNAc) [59]. To test whether any effects were dependent on the imported substrate, growth curves and transformation experiments were undertaken in the presence of these carbon source supplements. Few significant differences in growth rate were observed, except a small positive effect of GlcNAc on the growth of $RMV7_{wt}$ (Fig. S26). However, GlcNAc increased transformation efficiency ~10-fold in $RMV7_{wt}$ (Fig. 3A) and ~2-fold in $RMV7_{rare}$ (Fig. 3B). In both variants, this effect was dependent upon *manLMN*. This was confirmed through disrupting, and then restoring, *manL* alone (Fig. S27). Therefore, uptake of GlcNAc by ManLMN increased the transformability of RMV7.

_N_-acetylglucosamine activates competence through TfoX and Yjbk

This result suggested two parallels with the regulation of competence in gram-negative bacteria. Competence in *Vibrio cholerae* is induced GlcNAc disaccharides [60], thought to be generated from degradation of chitin [61]. This is mediated through the Transformation Factor X (TfoX) protein [62,63]. An orthologue of this protein (also called Sxy) is also central to regulation of competence in *Haemophilus influenzae* [64], in which it facilitates induction via 3',5'-cyclic AMP (cAMP) signalling [65,66]. Intracellular concentrations of 3',5'-cAMP rise in many proteobacteria when the primary glucose PTS transporter is inactive, as the accumulation of phosphorylated EIIA PTS subunits stimulate adenylate cyclase activity [67]. We tested whether analogues of either of these pathways existed in *S. pneumoniae*.

An orthologue of the N terminal domain of TfoX was annotated, but not described, in *S. pneumoniae* ATCC 700669 [68]. In RMV7, this gene (*tfoX*$_{Spn}$) is conserved in the same position, two genes upstream of the *comEA* competence operon (Fig. S28). The amino acid sequence corresponded to only the N-terminal region of the gram-negative orthologue, and was predicted to form a four-strand beta sheet flanked by alpha helices (Fig. S29). The gene could be both disrupted, and restored, in RMV7$_{rare}$. Transformation assays using a panel of sugars demonstrated the elevated induction of competence by GlcNAc in this variant was dependent on TfoX$_{Spn}$ (Fig. 3C).

A gene encoding a candidate adenylate cyclase, *yjbK*, was also identified in RMV7. This protein is predicted to have a β-barrel structure, as observed for orthologous enzymes synthesising 3',5'-cAMP (Fig. S30). The *yjbK* gene could be both disrupted, and restored, in RMV7$_{rare}$. Transformation assays with these genotypes also demonstrated the effect of GlcNAc on competence induction in RMV7$_{rare}$ also depended on YjbK (Fig. 3D). However, no 3',5'-cAMP signalling pathway is known in Firmicutes [69]. Therefore, an ELISA was used to compare the 3',5'-cAMP levels in RMV7$_{wt}$, RMV7$_{rare}$ *yjbK*::Janus and RMV7$_{rare}$ *yjbK* restored, relative to *Escherichia coli* DH5α. High levels of 3',5'-cAMP were found in fractions collected from both exponential and stationary phases of *E. coli.* However, the concentrations of 3',5'-cAMP detected in *S. pneumoniae* were close to the lower detection threshold, and unaffected by *yjbK* presence (Fig. 3E). Therefore, it is unlikely that YjbK's regulatory role is mediated through 3',5'-cAMP production.

To test whether TfoX and YjbK had a similar role in mediating the larger response to GlcNAc in RMV7$_{wt}$, both were also knocked out in this background. However, the results showed a different pattern (Fig. 3F). The disruption of *tfoX* had no effect on GlcNAc's ability to induce competence, whereas RMV7$_{wt}$ *yjbK*::Janus failed to activate competence unless GlcNAc or sialic acid were supplemented in the media. These results were consistent with ManLMN affecting competence through at least two pathways, one of which is dependent upon TfoX and YjbK. Disruption of these genes in the laboratory strain R6x was also found to reduce transformation efficiency, although detecting these effects required culturing in a low-sugar chemically-defined medium (see Methods; Fig. S31).

Variation in mobile element activity associated with cell stress

Although ManLMN had a role in regulating the competence machinery, it could not account for the large difference in transformability between RMV7$_{domi}$ and RMV7$_{rare}$ alone. MGEs have been found to strongly affect levels of competence in *S. pneumoniae* and other species

[16,70–72]. Therefore we tested the hypothesis that the increased activity of PRCI$_{dnaN}$ in RMV7$_{domi}$ may inhibit the competence of the host cell, in order to prevent the MGE being deleted through homologous recombination [17]. A candidate gene (IONPJBJN_00496) with the potential to limit homologous recombination encoded a protein similar to DinD (DNA damage-inducible protein D), which inhibits RecA activity in *E. coli* [73]. However, the RMV7$_{wt}$ IONPJBJN_00496::Janus mutant did not show any increase in transformability (Fig. 4A).

To test whether any other part of the PRCI might inhibit transformation, the entire element was removed, either with (RMV7$_{wt}$ PRCI$_{dnaN+att}$::Janus) or without (RMV7$_{wt}$ PRCI$_{dnaN}$::Janus) the flanking *att* sites. In both cases, a ~5-fold increase in transformation rates was observed (Fig. 4A). This implied the PRCI encoded an activity that inhibited the activation of the competence system. To identify where this was located within the MGE, four large mutations were generated within the PRCI: one removing the regulatory genes; one removing the regulatory genes and IONPJBJN_00496; one removing the replication genes; and one removing both the regulatory and replication genes. However, none of these mutations had such a large effect on transformation rates as the elimination of the entire element (Fig. 4A). This suggested the inhibition of competence induction was not the consequence of a single gene product, but instead the activity of the MGE itself.

A qRT-PCR assay was employed to test whether activation of PRCI$_{dnaN}$ could affect transformation through upregulating other stress signals within the cell. This found neither *ciaR* nor *manL* expression was altered when the PRCI was removed (Fig. 4C). However, transcription of the chaperone regulator *hrcA* was halved in the absence of the MGE, consistent with the higher expression of this gene in RMV7$_{domi}$ pre-CSP (Table S3). This indicated that the repression of competence by PRCI$_{dnaN}$ could occur indirectly, through MGE-driven elevation of the protein misfolding responses in the cell.

Repression of competence by conserved chaperone proteins

HrcA is encoded by a gene within the *dnaK* heat shock operon, and it autoregulates through repressing this chaperone-encoding gene cluster through binding the CIRCE motif [74]. In *S. pneumoniae*, HrcA binding of CIRCE is reduced at elevated temperatures, relieving its repression of the *dnaK* and *groEL* operons, enabling a heat shock response [75]. By contrast, Ca^{2+} ions facilitate HrcA-CIRCE motif binding, inhibiting chaperone expression [76]. Low Ca^{2+} concentrations and extreme temperatures, both of which reduce HrcA-CIRCE interactions, have been found to inhibit the induction of competence [77]. To test whether

HrcA mediated these effects on transformation efficiency, the *hrcA* gene was disrupted with a Janus cassette, then restored, in $RMV7_{rare}$.

Disruption of *hrcA* was confirmed to relieve repression of *dnaK* and *groL* expression through qRT-PCR, with no changes in the *ciaR*, demonstrating that any effects on transformation were not mediated through CiaRH (Fig. 5A). In the absence of Ca^{2+}, the *hrcA*::Janus mutant had a lower transformation efficiency than either $RMV7_{rare}$, or $RMV7_{rare}$ *hrcA* restored (Fig. 5B). This difference grew as the Ca^{2+} concentration increased, consistent with HrcA-CIRCE binding increasing transformation efficiency. These effects could be reproduced in *S. pneumoniae* R6x grown in a chemically-defined medium (Fig. S32). Unfortunately, the essential nature of the chaperones regulated by HrcA prevented any further identification of the specific mechanism by which competence was inhibited.

Replicating previous observations, a 40 °C heat shock reduced the transformation efficiency of both $RMV7_{rare}$ and $RMV7_{wt}$ (Fig. 5C). However, the heat shock increased the transformation efficiency of $RMV7_{rare}$ *hrcA*::Janus, demonstrating the competence machinery itself to be resilient to misfolding at elevated temperatures. Hence the effect of elevated temperature on transformation efficiency appears to be mediated through the decreased CIRCE-HrcA interactions. This cannot explain the increased transformation efficiency of the *hrcA*::Janus cells after heat shock, but the growth curves of these genotypes at 40 °C suggests the loss of *hrcA* has pleiotropic effects on pneumococcal physiology (Fig. S33).

To check that this regulation was independent of the ManLMN-mediated effects on competence, the transformation efficiency of the $RMV7_{rare}$ *hrcA*::Janus *manLMN*::*cat* double mutant was compared to that of the progenitor genotype, and the single mutants (Fig. 5D). This demonstrated an approximately five-fold decrease in transformation for each single mutant, and a ~25-fold reduction for the double mutant. This is consistent with HrcA and ManLMN both activating competence independently.

Discussion

The transformation efficiency of *S. pneumoniae* is highly heterogeneous across populations [9,10,78] and over the history of individual strains [8,79]. Yet the competence machinery itself is highly conserved across the species [2]. *S. pneumoniae* RMV7 is a representative of GPSC1 [38], which has diversified through homologous recombination exceptionally quickly, enabling vaccine escape and the acquisition of antibiotic resistance loci [7]. However, even within a single isolate of this strain, only a minority of the population were highly transformable *in vitro*. Phase variable systems enable such phenotypic heterogeneity to

persist within a near-isogenic population [27,32], resulting in a “bet hedging” strategy in RMV7 that mimics the heterogeneity in transformation efficiency of *Bacillus subtilis* cultures [33,80,81]. Although the competence state is regarded as a stress response analogous to the SOS response in other bacteria [47], this work suggests its activation requires a nutrient-rich milieu and an absence of abiotic stresses, such as elevated temperature. The alternative stress state, observed in $RMV7_{domi}$, upregulates the chaperones and proteases needed to cope with protein misfolding, and proteins thought to ameliorate cell wall disruption, without the protein synthesis and energy expenditure of competence induction. A mixture of stress-response strategies likely enables the pneumococcus to survive a greater diversity of environmental challenges.

This heterogeneity might be unexpected, given CSP’s role in synchronising competence induction across pneumococcal populations. The competence inhibitor CiaRH inhibits pheromone signalling through induction of five csRNAs, inhibiting intracellular CSP production [52,82,83], and the upregulation of HtrA, which degrades extracellular CSP [53]. If the pheromone were freely diffusing throughout a population, such inhibition of competence would be homogenous across cells. However, the downregulation of *manLMN* by CiaRH reduces cell’s ability to respond to a CSP stimulus, enabling heterogeneity within a single population.

Despite its major role in regulating the competence machinery, ManLMN was dispensable for efficient growth in rich media. This transporter is highly conserved across *S. pneumoniae* (Fig. S34; Table S4), and orthologues serve as the main glucose transporter across many Firmicutes, including other streptococci, *Lactococcus lactis* and *Listeria monocytogenes* [59]. In *S. pneumoniae*, it acts as the central metabolic regulator, with the ManL EIIAB component important in controlling the carbon catabolite preferences of *S. pneumoniae* cells [84]. Loss of ManLMN in *Streptococcus mutans* was associated with decreased biofilm formation and transformability [85], suggesting its regulatory role is likely to be shared across many streptococci.

While *S. pneumoniae* encodes a plethora of transporters [26], experiments in chemically-defined media suggest ManLMN may be the only effective means of importing GlcNAc [59]. Crucially, ManLMN’s role as the primary glucose importer means GlcNAc can act as an intracellular signal without being subject to catabolite repression. On importation, GlcNAc is converted to the cytotoxic compound GlcNAc-6-phosphate. This is typically used for cell wall biosynthesis in proliferating cells, but any excess must be converted to fructose-6-phosphate for glycolysis [86]. Hence GlcNAc is a uniquely informative signal of nutrient availability

relative to cell proliferation [87]. The increased availability of GlcNAc by *V. cholerae* has been suggested to signal the colonisation of a copepod's chitinous surface [88].

Analogously, GlcNAc is a major component of mucins covering the upper respiratory tract [89], although *S. pneumoniae* is an obligate commensal of the human nasopharynx, and likely derives little ecological information from GlcNAc's availability from host structures. An alternative interpretation is locally-elevated concentrations of GlcNAc may be indicative of lysis of bacterial cell walls, as suggested for streptomycetes [90]. As *V. cholerae* and *S. pneumoniae* both co-ordinate fratricide with their competence machinery [91], this would represent a parsimonious explanation for both using GlcNAc as a signal inducing competence.

These two pathogens also both regulate competence through quorum sensing (Fig. 6). The detection of quorum sensing and elevated GlcNAc concentrations would indicate high local densities of conspecific bacteria undergoing cell wall lysis, and therefore elevated concentrations of genomic DNA suitable for integration through homologous recombination. There are further similarities between the other components of the GlcNAc-signalling pathway. The N-terminal TfoX domain is present in many bacterial phyla (Fig. S35), while $YjbK_{Spn}$ belongs to a recently-defined subset of CYTH proteins [92,93] that contains orthologues of unknown function in *B. subtilis* [94] and *Staphylococcus aureus* [69]. Both TfoX and YjbK are conserved within *S. pneumoniae* itself (Fig S34), although *tfoX* is absent from ~12% of isolates [95]. Neither is strongly upregulated by CSP, making it difficult to identify their regulatory functions through gene expression studies alone (Fig. S36). In gram-negative bacteria, TfoX affects transcription [71], although it is unclear whether the shorter orthologue found in gram-positive bacteria would be able fulfil the same role. The function of YjbK is even less clear, as 3,5'-cAMP is undetectable in Firmicutes, to whom it is toxic [69]. This protein may be considered a candidate for generating an alternative signalling molecule.

It seems likely there is at least one more mechanism by which GlcNAc and ManLMN affect the competence machinery, as the promotion of transformation by GlcNAc in $RMV7_{wt}$ was independent of these proteins. Such upregulation by sugar import contrasts with the activation of transformation by nucleotide starvation in *H. influenzae* [96], which led to the hypothesis that the competence machinery primary evolved to scavenge metabolites from extracellular DNA [97,98]. This interspecies difference likely accounts for *H. influenzae* transformation being inducible in stationary phase, whereas *S. pneumoniae* transformation *in vitro* requires cultures in the early exponential phase [99].

In vitro transformation of *S. pneumoniae* also routinely involves the addition of $CaCl_2$. This was previously suggested to aid the translocation of DNA molecules across the plasma membrane [100]. However, we found this effect was dependent upon HrcA, the binding of which to CIRCE motifs is promoted by higher Ca^{2+} concentrations [76]. Both the deletion of *hrcA*, and the increased expression in $RMV7_{domi}$ associated with PRCI-associated stress, were associated with reduced transformation efficiency. This suggests the activation of competence is mediated by HrcA-CIRCE complexes. Any involvement of the chaperones regulated by HrcA could not be tested in this genetic study. In other species, orthologues of the chaperones regulated by HrcA-CIRCE have been found to block sigma factor switching, as required for the activation of late competence genes [101]. However, given the chaperones did not appear to significantly change their expression in response to alterations in PRCI replication, HrcA may affect competence genes through a chaperone-independent route. The effect of this regulator on transformation was most sensitive to changes in Ca^{2+} concentration in the 0-5 mM range, spanning the relevant concentrations of the ion in mucus [102] and interstitial fluid, which rise on cell necrosis and inflammation [103]. Hence HrcA likely represents a physiologically-relevant competence regulator.

The level of complexity of competence regulation in *S. pneumoniae* is similar to that in *Bacillus subtilis* [104], as well as *Vibrio cholerae* [105] and other gram-negative bacteria [106]. Such signalling networks may be the product of adaptive evolution, selecting for the integration of many external cues to trigger competence only under optimal conditions. However, this would require that *S. pneumoniae* benefits from transformation when GlcNAc is abundant, whereas *H. influenzae* benefits during starvation conditions, when both species are frequently isolated within the same nasopharynx [107]. Perhaps tellingly, the regulatory roles of ManLMN and HrcA are moonlighting functions, secondary physiological activities separate from their primary physiological role [108]: the transporter did not affect growth in rich media (Fig. 3), and the chaperone regulator was not necessary to enable competence at higher temperatures (Fig. 5).

Their activation of competence may instead be a consequence of gene-level selection. Individual cells, and therefore all genes, must benefit from transformation. Yet each gene also suffers a cost, as it may be replaced by an imported allele through homologous recombination. The costs to variable loci, such as MGEs, is higher. This is because they are likely to differ between donor and recipient bacteria, and therefore are most likely to suffer replacement with a different sequence [17]. Hence they may evolve to selfishly inhibit transformation. However, the cost to conserved loci is lower, as their sequences are unlikely

to be altered by transformation, even if they are affected by homologous recombination. If synchronising competence across populations is important, then conserved regulators (ManLMN and HrcA in *S. pneumoniae*; the Catabolite Repression Protein in *H. influenzae*) may acquire secondary competence-promoting activities. This is because purifying selection for their primary function will limit population-wide sequence diversity, meaning they experience the greatest net benefit from transformation. These moonlighting functions may be necessary to counteract the influence of variable loci under selection to reduce transformation [17,98]. Such conflicting gene-level pressures may explain why so many loci regulate the competence machinery, and why transformation efficiency is so variable across and within species [109].

Methods and materials

Cell culture and mutagenesis

Genotypes used in this study are described in Table S5. Unless otherwise stated, encapsulated *S. pneumoniae* were cultured statically at 35 °C with 5% CO_2 in 10 mL of a mixed liquid media, consisting of a 2:3 ratio of Todd-Hewitt media with 0.5% yeast extract (Sigma-Aldrich), and Brain-Heart Infusion media (Sigma-Aldrich). Transformation experiments with *S. pneumoniae* R6x derivatives used a chemically-defined medium, consisting of disodium β-glycerophosphate (20 g L^{-1}; Sigma), sodium pyruvate (0.1 g L^{-1}; Fluorochem), choline (0.001 g L^{-1}; Alfa Aesar), cysteine (0.4 g L^{-1}; Tokyo Chemical Industry UK), glucose (3.8 mM; Sigma) and galactose (12 mM; Sigma). Carbohydrates were added as a supplement to liquid media at a final concentration of 33 mM, unless otherwise specified.

Culturing on solid media used Todd-Hewitt media supplemented with 0.5% yeast extract. For the generation of mutants, 0.8-1 kb PCR fragments of regions flanking the gene of interest were amplified with added *Bam*HI and *Eco*RV restriction enzyme sites. Oligonucleotide sequences are listed in Table S6. PCR products were digested with the appropriate restriction enzymes (Promega) at 35 °C for 2-4 hours, and then ligated to the chosen antibiotic markers using T4 DNA ligase (Invitrogen). Ligations were used as templates for further amplifications using PCR. The amplified PCR constructs were then transformed into the cells.

Transformation assays

One millilitre of the bacterial culture was collected at an OD_{600} between 0.15-0.25. Cells were then incubated for 2 hours at 35 °C with 5 μl of 500 mM $CaCl_2$ (Sigma), 250 ng of competence stimulating peptide 1 (CSP-1; Cambridge Bioscience Ltd) and 100 ng of the purified *rpoB* gene, containing a SNP which conferred resistance to rifampicin [110]. After two hours of incubation at 35 °C, a volume of between 1 and 100 μl of the transformed culture was spread on an agar plate supplemented with 4 μg mL^{-1} of rifampicin (Fisher Scientific). For precise quantification of transformation frequencies, 1 μl of a 10^3-fold dilution of the same culture was spread on a non-selective plate in parallel. Colonies were counted after 24 hours of incubation at 35 °C with 5% CO_2. For other transformations, the same conditions were used, and cells were incubated on agar plates supplemented with the appropriate selective antibiotics: kanamycin (Sigma) at 600 μg mL^{-1}, or chloramphenicol (Sigma) at 4 μg mL^{-1}.

A logistic curve was used to analyse the relationship between transformation efficiency, *t*, and the volume of 25 mM $CaCl_2$ added to the 1 mL of cells, *v*. The fitted function was:

$$t = \frac{a}{1 + e^{-b(v-c)}}$$

The values of the variables *a*, *b* and *c* were estimated using the Levenberg-Marquardt nonlinear least-squares algorithm in the minpack R package [111], using the starting values of 10 (increased to 500 for *S. pneumoniae* R6x), 1 and 0.5, respectively. The confidence intervals were calculated through refitting the function to 999 bootstrapped samples using the car R package [112].

Biofilm assays and growth curves

To measure growth curves, $2x10^4$ cells from titrated frozen stocks were grown in mixed liquid media in 96-well microtiter plates at 35°C with 5% CO_2 for 20 h. Measurements of OD_{600} were taken at 30 minutes intervals over 16 hours using the FLUOstar Omega microplate reader (BMG LABTECH). Three replicate wells were used for each tested genotype in each experiment. For measuring biofilms, at the end of the growth curve incubation, the microtitre plate was submerged in water and dried for 10 minutes. Then 150 μl of a freshly-diluted 0.1% crystal violet solution (Scientific Laboratory Supplies) was added to each well and incubated for 30 minutes at room temperature. Each well was then washed by repeatedly submerging the plate in water to remove excess crystal violet. The plate was incubated at room temperature in an inverted position for four hours. Subsequently, 150 μl of 30% acetic acid (Honeywell) was added to each well. Wells were scanned using a FLUOstar Omega plate reader at OD_{550}.

Preparations of RNA samples and quantitative PCR

Three replicate cultures of RMV7 tvr_{domi}::Janus and RMV7 tvr_{rare}::Janus were grown in 25 mL of mixed liquid media until they reached an OD_{600} of 0.15. A 5 mL sample of bacterial cells was collected and 5 μL 250 ng mL^{-1} CSP1 was added to the remaining culture. Further 5 mL samples were taken from each culture 10 and 20 min post-CSP addition. Each sample was immediately treated with 10 mL RNAprotect (Qiagen) and incubated at room temperature for 5 min. Cell were then pelleted by centrifugation at 4,000 rpm for 10 min. RNA was extracted from the washed pellets using the SV Total RNA Isolation System (Promega) according to the manufacturer's instructions. The extracted RNA was used for RNA sequencing or qRT-PCR.

All qRT-PCR experiments were conducted as described previously [34], using 0.2 μg of RNA to generate cDNA with the First-Strand III cDNA synthesis kit (Invitrogen). Reactions used the PowerUp™ SYBR™ Green Master Mix (Thermo Fisher) and the QuantStudio™ 7 Flex Real-Time PCR System (Applied Biosystems).

Analysis of RNA-seq data

The set of genes used for expression analysis were the 2,088 protein coding sequences annotated on the *S. pneumoniae* $RMV7_{domi}$ genome (accession code OV904788), and the 81 non-coding RNAs predicted by infernal version 1.1.2 [113] using the Rfam database [114]. RNA-seq reads were aligned to these sequences using kallisto version 0.46.2 [115], using default settings and 100 bootstraps. Differential gene expression analysis used sleuth version 0.30 [116]. Wald tests were conducted to compare the pre-CSP samples for $RMV7_{domi}$ and $RMV7_{rare}$, and to compare the 10 min and 20 min post-CSP samples for each genotype to the corresponding pre-CSP samples. Visualisation and plotting of data used the genoplotR [117], circlize [118], cowplot [119], ggpubr [120] and tidyverse [121] packages.

Quantification of 3',5'-cAMP production

Quantification of 3',5'-cAMP used the Cyclic AMP XP® assay kit (Cell Signalling Technology). *S. pneumoniae* cultures were grown statically as described above, and samples were harvested in the exponential (OD_{600} of 0.2) and stationary (OD_{600} of 0.5) phases. *E. coli* DH5α was grown in 25 mL Luria-Bertani media (Sigma) with 250 rpm at 37 °C, and samples were harvested in the exponential (OD_{600} of 0.3) and stationary (OD_{600} of 1.0) phases. Cells were pelleted through centrifugation at 4,000 rpm for 10 min. The cell pellets were re-suspended in 850 μL of the kit's lysis buffer with 150 μl of lysozyme (10 mg mL^{-1}), followed by incubation at 35 °C for 20 minutes. Cells were pelleted by centrifugation at 12,000 rpm for 5 minutes. The supernatants were collected, and the cell pellets were suspended in 400 μl phosphate-buffered saline. The protein concentrations of all the collected samples were adjusted to 400 μg mL^{-1} using the Qubit protein broad range assay kits (Qiagen). The kit ELISA plates were prepared according to the manufacturer's protocols, and a 50 μl sample of supernatant from each tested culture was loaded in each well. The OD_{450} was measured using the FLUOstar Omega plate reader. Concentrations of 3',5'-cAMP were calculated using a standard curve, according to the manufacturer's instructions.

Analysis of motif distribution and sequence diversity

The overall distribution of methylation motifs was calculated using DistAMo [122]. The calculation of distances between coding sequences and motifs used biopython.

The diversity of genes encoding regulatory proteins within *S. pneumoniae* used previously-calculated pairwise genetic distances [123]. Only genes found in at least half the isolates in the collection were used, to avoid genes for which only a small sample size was available.

Proteins containing the TfoX N-terminal domain were identified using EMBL SMART [124]. These amino acid sequences were aligned with MAFFT [125], and a phylogeny generated with Fasttree2 [126] using default settings. Proteins were assigned to Families using the NCBI Taxonomy [127].

References

1. Griffith F. The Significance of Pneumococcal Types. J Hyg (Lond). 1928;27: 113–159. doi:10.1017/S0022172400031879
2. Johnston C, Campo N, Bergé MJ, Polard P, Claverys JP. *Streptococcus pneumoniae*, le transformiste. Trends Microbiol. 2014;22: 113–9. doi:10.1016/j.tim.2014.01.002
3. Dowson CG, Hutchison A, Brannigan JA, George RC, Hansman D, Linares J, et al. Horizontal transfer of penicillin-binding protein genes in penicillin-resistant clinical isolates of *Streptococcus pneumoniae*. Proc Natl Acad Sci. 1989;86: 8842–8846. doi:10.1073/pnas.86.22.8842
4. Sibold C, Henrichsen J, Konig A, Martin C, Chalkley L, Hakenbeck R. Mosaic *pbpX* genes of major clones of penicillin-resistant *Streptococcus pneumoniae* have evolved from *pbpX* genes of a penicillin-sensitive *Streptococcus oralis*. Mol Microbiol. 1994/06/01. 1994;12: 1013–1023.
5. Dowson CG, Coffey TJ, Kell C, Whiley RA. Evolution of penicillin resistance in *Streptococcus pneumoniae*; the role of *Streptococcus mitis* in the formation of a low affinity PBP2B in *S. pneumoniae*. Mol Microbiol. 1993/08/01. 1993;9: 635–643. doi:10.1111/j.1365-2958.1993.tb01723.x
6. D'Aeth JC, van der Linden MPG, McGee L, de Lencastre H, Turner P, Song J-H, et al. The role of interspecies recombination in the evolution of antibiotic-resistant pneumococci. Elife. 2021;10: e67113. Available: https://doi.org/10.7554/eLife.67113
7. Croucher NJ, Chewapreecha C, Hanage WP, Harris SR, McGee L, van der Linden M, et al. Evidence for soft selective sweeps in the evolution of pneumococcal multidrug resistance and vaccine escape. Genome Biol Evol. 2014;6: 1589–1602. doi:10.1093/gbe/evu120
8. Golubchik T, Brueggemann AB, Street T, Gertz RE, Spencer CCA, Ho T, et al. Pneumococcal genome sequencing tracks a vaccine escape variant formed through a multi-fragment recombination event. Nat Genet. 2012;44: 352–355. doi:10.1038/ng.1072
9. Croucher NJ, Finkelstein JA, Pelton SI, Mitchell PK, Lee GM, Parkhill J, et al. Population genomics of post-vaccine changes in pneumococcal epidemiology. Nat Genet. 2013;45: 656–663. doi:10.1038/ng.2625
10. Croucher NJ, Mitchell AM, Gould KA, Inverarity D, Barquist L, Feltwell T, et al. Dominant role of nucleotide substitution in the diversification of serotype 3 pneumococci over decades and during a single infection. PLoS Genet. 2013;9: e1003868. doi:10.1371/journal.pgen.1003868
11. Evans BA, Rozen DE. Significant variation in transformation frequency in

Streptococcus pneumoniae. ISME J. 2013;7: 791–799. doi:10.1038/ismej.2012.170

12. Hsieh YC, Wang JT, Lee W Sen, Hsueh PR, Shao PL, Chang LY, et al. Serotype competence and penicillin resistance in *Streptococcus pneumoniae*. Emerg Infect Dis. 2006;12: 1709–1714. doi:10.3201/eid1211.060414
13. Joloba ML, Kidenya BR, Kateete DP, Katabazi FA, Muwanguzi JK, Asiimwe BB, et al. Comparison of transformation frequencies among selected *Streptococcus pneumoniae* serotypes. Int J Antimicrob Agents. 2010;36: 124–128. doi:10.1016/j.ijantimicag.2010.03.024
14. Yother J, McDaniel LS, Briles DE. Transformation of encapsulated *Streptococcus pneumoniae*. J Bacteriol. 1986;168: 1463–1465. doi:10.1128/jb.168.3.1463-1465.1986
15. Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, et al. Rapid pneumococcal evolution in response to clinical interventions. Science (80-). 2011/01/29. 2011;331: 430–434. doi:10.1126/science.1198545
16. Croucher NJ, Hanage WP, Harris SR, McGee L, van der Linden M, de Lencastre H, et al. Variable recombination dynamics during the emergence, transmission and “disarming” of a multidrug-resistant pneumococcal clone. BMC Biol. 2014;12: 49. doi:10.1186/1741-7007-12-49
17. Croucher NJ, Mostowy R, Wymant C, Turner P, Bentley SD, Fraser C. Horizontal DNA Transfer Mechanisms of Bacteria as Weapons of Intragenomic Conflict. Barton NH, editor. PLOS Biol. 2016;14: e1002394. doi:10.1371/journal.pbio.1002394
18. Johnston C, Martin B, Fichant G, Polard P, Claverys JP. Bacterial transformation: Distribution, shared mechanisms and divergent control. Nat Rev Microbiol. 2014;12: 181–196. doi:10.1038/nrmicro3199
19. Håvarstein LS, Coomaraswamy G, Morrison DA. An unmodified heptadecapeptide pheromone induces competence for genetic transformation in *Streptococcus pneumoniae*. Proc Natl Acad Sci. 1995;92: 11140 LP – 11144. doi:10.1073/pnas.92.24.11140
20. Alloing G, Martin B, Granadel C, Claverys J-P. Development of competence in *Streptococcus pneumoniae*: pheromone autoinduction and control of quorum sensing by the oligopeptide permease. Mol Microbiol. 1998;29: 75–83. doi:https://doi.org/10.1046/j.1365-2958.1998.00904.x
21. S. LM, A. MD. Identification of a New Regulator in *Streptococcus pneumoniae* Linking Quorum Sensing to Competence for Genetic Transformation. J Bacteriol. 1999;181: 5004–5016. doi:10.1128/JB.181.16.5004-5016.1999
22. Campbell EA, Choi SY, Masure HR. A competence regulon in *Streptococcus pneumoniae* revealed by genomic analysis. Mol Microbiol. 1998;27: 929–939.

doi:10.1046/j.1365-2958.1998.00737.x
23. Pestova E V, Morrison DA. Isolation and characterization of three *Streptococcus pneumoniae* transformation-specific loci by use of a *lacZ* reporter insertion vector. J Bacteriol. 1998;180: 2701–2710. doi:10.1128/JB.180.10.2701-2710.1998
24. Kyoo SC, A. MD. Two Distinct Functions of ComW in Stabilization and Activation of the Alternative Sigma Factor ComX in *Streptococcus pneumoniae*. J Bacteriol. 2005;187: 3052–3061. doi:10.1128/JB.187.9.3052-3061.2005
25. Weiser JN, Kapoor M. Effect of intrastrain variation in the amount of capsular polysaccharide on genetic transformation of *Streptococcus pneumoniae*: Implications for virulence studies of encapsulated strains. Infect Immun. 1999;67: 3690–3692.
26. Tettelin H, Nelson KE, Paulsen IT, Eisen JA, Read TD, Peterson S, et al. Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. Science (80-). 2001;293: 498–506. doi:10.1126/science.1061217
27. Manso AS, Chai MH, Atack JM, Furi L, De Ste Croix M, Haigh R, et al. A random six-phase switch regulates pneumococcal virulence via global epigenetic changes. Nat Commun. 2014;5: 5055. doi:10.1038/ncomms6055
28. Croucher NJ, Coupland PG, Stevenson AE, Callendrello A, Bentley SD, Hanage WP. Diversification of bacterial genome content through distinct mechanisms over different timescales. Nat Commun. 2014;5: 5471.
29. Li JW, Li J, Wang J, Li C, Zhang JR. Molecular mechanisms of *hsdS* inversions in the *cod* locus of *Streptococcus pneumoniae*. J Bacteriol. 2019;201: e00581-18. doi:10.1128/JB.00581-18
30. Li J, Li JW, Feng Z, Wang J, An H, Liu Y, et al. Epigenetic Switch Driven by DNA Inversions Dictates Phase Variation in *Streptococcus pneumoniae*. PLoS Pathog. 2016;12: e1005762. doi:10.1371/journal.ppat.1005762
31. Wang J, Li JW, Li J, Huang Y, Wang S, Zhang JR. Regulation of pneumococcal epigenetic and colony phases by multiple two-component regulatory systems. PLoS Pathog. 2020;16: e1008417. doi:10.1371/journal.ppat.1008417
32. De Ste Croix M, Vacca I, Kwun MJ, Ralph JD, Bentley SD, Haigh R, et al. Phase-variable methylation and epigenetic regulation by type I restriction-modification systems. FEMS Microbiol Rev. 2017;41: S3–S15. doi:10.1093/femsre/fux025
33. Veening J-W, Smits WK, Kuipers OP. Bistability, epigenetics, and bet-hedging in bacteria. Annu Rev Microbiol. 2008;62: 193–210. doi:10.1146/annurev.micro.62.081307.163002
34. Kwun MJ, Oggioni MR, De Ste Croix M, Bentley SD, Croucher NJ. Excision-reintegration at a pneumococcal phase-variable restriction-modification locus drives within- and between-strain epigenetic differentiation and inhibits gene acquisition.

Nucleic Acids Res. 2018;46: 11438–11453. doi:10.1093/nar/gky906

35. Kwun MJ, Oggioni MR, Bentley SD, Fraser C, Croucher NJ. Synergistic Activity of Mobile Genetic Element Defences in *Streptococcus pneumoniae*. Genes (Basel). 2019;10: 707.

36. Slager J, Aprianto R, Veening JW. Refining the pneumococcal competence regulon by RNA sequencing. J Bacteriol. 2019;201: e00780-18. doi:10.1128/JB.00780-18

37. Sung CK, Li H, Claverys JP, Morrison DA. An *rpsL* cassette, Janus, for gene replacement through negative selection in *Streptococcus pneumoniae*. Appl Environ Microbiol. 2001;67: 5190–5196.

38. Gladstone RA, Lo SW, Lees JA, Croucher NJ, van Tonder AJ, Corander J, et al. International genomic definition of pneumococcal lineages, to contextualise disease, antibiotic resistance and vaccine impact. EBioMedicine. 2019;43: 338–346. doi:10.1016/j.ebiom.2019.04.021

39. Trappetti C, Kadioglu A, Carter M, Hayre J, Iannelli F, Pozzi G, et al. Sialic Acid: A Preventable Signal for Pneumococcal Biofilm Formation, Colonization, and Invasion of the Host. J Infect Dis. 2009;199: 1497–1505. doi:10.1086/598483

40. Li Y, Croucher NJ, Thompson CM, Trzciński K, Hanage WP, Lipsitch M. Identification of pneumococcal colonization determinants in the stringent response pathway facilitated by genomic diversity. BMC Genomics. 2015;16: 369. doi:10.1186/s12864-015-1573-6

41. Lamarche MG, Wanner BL, Crépin S, Harel J. The phosphate regulon and bacterial virulence: a regulatory network connecting phosphate homeostasis and pathogenesis. FEMS Microbiol Rev. 2008;32: 461–473. doi:10.1111/j.1574-6976.2008.00101.x

42. Aggarwal SD, Eutsey R, West-Roberts J, Domenech A, Xu W, Abdullah IT, et al. Function of BriC peptide in the pneumococcal competence and virulence portfolio. PLOS Pathog. 2018;14: e1007328. Available: https://doi.org/10.1371/journal.ppat.1007328

43. Peterson SN, Sung CK, Cline R, Desai B V, Snesrud EC, Luo P, et al. Identification of competence pheromone responsive genes in *Streptococcus pneumoniae* by use of DNA microarrays. Mol Microbiol. 2004;51: 1051–1070. doi:https://doi.org/10.1046/j.1365-2958.2003.03907.x

44. Majchrzykiewicz JA, Kuipers OP, Bijlsma JJE. Generic and specific adaptive responses of *Streptococcus pneumoniae* to challenge with three distinct antimicrobial peptides, bacitracin, LL-37, and nisin. Antimicrob Agents Chemother. 2009/11/16. 2010;54: 440–451. doi:10.1128/AAC.00769-09

45. Johnston CHG, Soulet A-L, Bergé M, Prudhomme M, De Lemos D, Polard P. The

alternative sigma factor σ_X mediates competence shut-off at the cell pole in Streptococcus pneumoniae. Blokesch M, Storz G, Blokesch M, Dubnau D, editors. Elife. 2020;9: e62907. doi:10.7554/eLife.62907

46. Weiser JN, Bae D, Epino H, Gordon SB, Kapoor M, A. ZL, et al. Changes in Availability of Oxygen Accentuate Differences in Capsular Polysaccharide Expression by Phenotypic Variants and Clinical Isolates of <i>Streptococcus pneumoniae<./i>. Infect Immun. 2001;69: 5430–5439. doi:10.1128/IAI.69.9.5430-5439.2001

47. Claverys J-P, Prudhomme M, Martin B. Induction of Competence Regulons as a General Response to Stress in Gram-Positive Bacteria. Annu Rev Microbiol. 2006;60: 451–475. doi:10.1146/annurev.micro.60.080805.142139

48. Hemsley C, Joyce E, Hava DL, Kawale A, Camilli A. MgrA, an Orthologue of Mga, Acts as a Transcriptional Repressor of the Genes within the rlrA Pathogenicity Islet in *Streptococcus pneumoniae*. J Bacteriol. 2003;185: 6640–6647. doi:10.1128/JB.185.22.6640-6647.2003

49. Solano-Collado V, Espinosa M, Bravo A. Activator role of the pneumococcal *mga*-like virulence transcriptional regulator. J Bacteriol. 2012;194: 197–4207. doi:10.1128/JB.00536-12

50. Martínez-Rubio R, Quiles-Puchalt N, Martí M, Humphrey S, Ram G, Smyth D, et al. Phage-inducible islands in the Gram-positive cocci. ISME J. 2017;11: 1029–1042. doi:10.1038/ismej.2016.163

51. Guenzi E, Gasc A -M, Sicard MA, Hakenbeck R. A two-component signal-transducing system is involved in competence and penicillin susceptibility in laboratory mutants of *Streptococcus pneumoniae*. Mol Microbiol. 1994;12: 505–515. doi:10.1111/j.1365-2958.1994.tb01038.x

52. Halfmann A, Kovács M, Hakenbeck R, Brückner R. Identification of the genes directly controlled by the response regulator CiaR in *Streptococcus pneumoniae*: Five out of 15 promoters drive expression of small non-coding RNAs. Mol Microbiol. 2007;66: 110–126. doi:10.1111/j.1365-2958.2007.05900.x

53. Cassone M, Gagne AL, Spruce LA, Seeholzer SH, Sebert ME. The HtrA protease from *Streptococcus pneumoniae* digests both denatured proteins and the competence-stimulating peptide. J Biol Chem. 2012;287: 38449–38459. doi:10.1074/jbc.M112.391482

54. Sebert ME, Palmer LM, Rosenberg M, Weiser JN. Microarray-based identification of *htrA*, a *Streptococcus pneumoniae* gene that is regulated by the CiaRH two-component system and contributes to nasopharyngeal colonization. Infect Immun. 2002;70: 4059–4067. doi:10.1128/IAI.70.8.4059-4067.2002

55. Stevens KE, Chang D, Zwack EE, Sebert ME. Competence in *Streptococcus*

pneumoniae is regulated by the rate of ribosomal decoding errors. MBio. 2011;2: e00071-11. doi:10.1128/mBio.00071-11

56. Dagkessamanskaia A, Moscoso M, Hénard V, Guiral S, Overweg K, Reuter M, et al. Interconnection of competence, stress and CiaR regulons in *Streptococcus pneumoniae*: Competence triggers stationary phase autolysis of *ciaR* mutant cells. Mol Microbiol. 2004;51: 1071–1086. doi:10.1111/j.1365-2958.2003.03892.x

57. Mascher T, Zähner D, Merai M, Balmelle N, De Saizieu AB, Hakenbeck R. The *Streptococcus pneumoniae cia* regulon: CiaR target sites and transcription profile analysis. J Bacteriol. 2003;185: 60–70. doi:10.1128/JB.185.1.60-70.2003

58. Gómez-Mejia A, Gámez G, Hammerschmidt S. *Streptococcus pneumoniae* two-component regulatory systems: The interplay of the pneumococcus with its environment. Int J Med Microbiol. 2018;308: 722–737. doi:10.1016/j.ijmm.2017.11.012

59. Bidossi A, Mulas L, Decorosi F, Colomba L, Ricci S, Pozzi G, et al. A functional genomics approach to establish the complement of carbohydrate transporters in *Streptococcus pneumoniae*. PLoS One. 2012;7: e33320. doi:10.1371/journal.pone.0033320

60. Yamamoto S, Morita M, Izumiya H, Watanabe H. Chitin disaccharide (GlcNAc)2 induces natural competence in *Vibrio cholerae* through transcriptional and translational activation of a positive regulatory gene $tfoX_{VC}$. Gene. 2010;457: 42–49. doi:10.1016/j.gene.2010.03.003

61. Meibom KL, Blokesch M, Dolganov NA, Wu CY, Schoolnik GK. Chitin induces natural competence in *Vibrio cholerae*. Science (80-). 2005;310: 1824–1827. doi:10.1126/science.1120096

62. Lo Scrudato M, Blokesch M. A transcriptional regulator linking quorum sensing and chitin induction to render *Vibrio cholerae* naturally transformable. Nucleic Acids Res. 2013;41: 3644–3658. doi:10.1093/nar/gkt041

63. Yamamoto S, Izumiya H, Mitobe J, Morita M, Arakawa E, Ohnishi M, et al. Identification of a chitin-induced small RNA that regulates translation of the *tfoX* gene, encoding a positive regulator of natural competence in *Vibrio cholerae*. J Bacteriol. 2011;193: 1953–1965. doi:10.1128/JB.01340-10

64. Williams PM, Bannister LA, Redfield RJ. The *Haemophilus influenzae sxy*-1 mutation is in a newly identified gene essential for competence. J Bacteriol. 1994;176: 6789–6794. doi:10.1128/jb.176.22.6789-6794.1994

65. Chandler MS. The gene encoding cAMP receptor protein is required for competence development in *Haemophilus influenzae* Rd. Proc Natl Acad Sci. 1992;89: 1626 LP – 1630. doi:10.1073/pnas.89.5.1626

66. Dorocicz IR, Williams PM, Redfield RJ. The *Haemophilus influenzae* adenylate cyclase gene: cloning, sequence, and essential role in competence. J Bacteriol. 1993;175: 7142–7149. doi:10.1128/jb.175.22.7142-7149.1993
67. Wu R, Zhao M, Li J, Gao H, Kan B, Liang W. Direct regulation of the natural competence regulator gene *tfoX* by cyclic AMP (cAMP) and cAMP receptor protein (CRP) in Vibrios. Sci Rep. 2015;5: 14921. doi:10.1038/srep14921
68. Croucher NJ, Walker D, Romero P, Lennard N, Paterson GK, Bason NC, et al. Role of conjugative elements in the evolution of the multidrug-resistant pandemic clone *Streptococcus pneumoniae*^{Spain23F} ST81. J Bacteriol. 2008/12/31. 2009;191: 1480–1489. doi:10.1128/jb.01343-08
69. Zhang Y, Agrebi R, Bellows LE, Collet J-F, Kaever V, Gründling A. Evolutionary Adaptation of the Essential tRNA Methyltransferase TrmD to the Signaling Molecule 3',5'-cAMP in Bacteria. J Biol Chem. 2017;292: 313–327. doi:10.1074/jbc.M116.758896
70. Durieux I, Ginevra C, Attaiech L, Picq K, Juan P-A, Jarraud S, et al. Diverse conjugative elements silence natural transformation in *Legionella* species. Proc Natl Acad Sci. 2019;116: 18613 LP – 18618. doi:10.1073/pnas.1909374116
71. Stutzmann S, Blokesch M. Comparison of chitin-induced natural transformation in pandemic *Vibrio cholerae* O1 El Tor strains. Environ Microbiol. 2020;22: 4149–4166. doi:https://doi.org/10.1111/1462-2920.15214
72. Dalia AB, Seed KD, Calderwood SB, Camilli A. A globally distributed mobile genetic element inhibits natural transformation of *Vibrio cholerae*. Proc Natl Acad Sci. 2015;112: 10485–10490. doi:10.1073/pnas.1509097112
73. Uranga LA, Balise VD, Benally C V, Grey A, Lusetti SL. The *Escherichia coli* DinD protein modulates RecA activity by inhibiting postsynaptic RecA filaments. J Biol Chem. 2011;286: 29480–29491. doi:10.1074/jbc.M111.245373
74. Zuber U, Schumann W. CIRCE, a novel heat shock element involved in regulation of heat shock operon *dnaK* of *Bacillus subtilis*. J Bacteriol. 1994;176: 1359–1363. doi:10.1128/jb.176.5.1359-1363.1994
75. Kim SN, Bae YG, Rhee DK. Dual regulation of *dnaK* and *groE* operons by HrcA and Ca^{++} in *Streptococcus pneumoniae*. Arch Pharm Res. 2008;31: 462–467. doi:10.1007/s12272-001-1179-4
76. Kwon HY, Kim SN, Pyo SN, Rhee DK. Ca^{2+}-dependent expression of the CIRCE regulon in *Streptococcus pneumoniae*. Mol Microbiol. 2005;55: 456 – 468. doi:10.1111/j.1365-2958.2004.04416.x
77. Fox MS, Hotchkiss RD. Initiation of bacterial transformation. Nature. 1957;179: 1322–1325. doi:10.1038/1791322a0

78. Chewapreecha C, Harris SR, Croucher NJ, Turner C, Marttinen P, Cheng L, et al. Dense genomic sampling identifies highways of pneumococcal recombination. Nat Genet. 2014;46: 305–309. doi:10.1038/ng.2895
79. Mostowy R, Croucher NJ, Hanage WP, Harris SR, Bentley S, Fraser C. Heterogeneity in the Frequency and Characteristics of Homologous Recombination in Pneumococcal Evolution. PLoS Genet. 2014;10: e1004300. doi:10.1371/journal.pgen.1004300
80. Süel GM, Garcia-Ojalvo J, Liberman LM, Elowitz MB. An excitable gene regulatory circuit induces transient cellular differentiation. Nature. 2006;440: 545–550. doi:10.1038/nature04588
81. Maamar H, Raj A, Dubnau D. Noise in gene expression determines cell fate in *Bacillus subtilis*. Science. 2007;317: 526–529. doi:10.1126/science.1140818
82. Laux A, Sexauer A, Sivaselvarajah D, Kaysen A, Brückner R. Control of competence by related non-coding csRNAs in *Streptococcus pneumoniae* R6. Front Genet. 2015;6: 246. doi:10.3389/fgene.2015.00246
83. Schnorpfeil A, Kranz M, Kovács M, Kirsch C, Gartmann J, Brunner I, et al. Target evaluation of the non-coding csRNAs reveals a link of the two-component regulatory system CiaRH to competence control in *Streptococcus pneumoniae* R6. Mol Microbiol. 2013;89: 334–349. doi:10.1111/mmi.12277
84. Fleming E, Camilli A. ManLMN is a glucose transporter and central metabolic regulator in *Streptococcus pneumoniae*. Mol Microbiol. 2016;102: 467–487. doi:https://doi.org/10.1111/mmi.13473
85. Abranches J, Candella MM, Wen ZT, Baker H V., Burne RA. Different Roles of $EIIAB^{Man}$ and EII^{Glc} in Regulation of Energy Metabolism, Biofilm Development, and Competence in *Streptococcus mutans*. J Bacteriol. 2006;188: 3748–3756. doi:10.1128/JB.00169-06
86. Kawada-Matsuo M, Mazda Y, Oogai Y, Kajiya M, Kawai T, Yamada S, et al. GlmS and NagB Regulate Amino Sugar Metabolism in Opposing Directions and Affect *Streptococcus mutans* Virulence. PLoS One. 2012;7: e33382. Available: https://doi.org/10.1371/journal.pone.0033382
87. Komatsuzawa H, Fujiwara T, Nishi H, Yamada S, Ohara M, McCallum N, et al. The gate controlling cell wall synthesis in *Staphylococcus aureus*. Mol Microbiol. 2004;53: 1221–1231. doi:10.1111/j.1365-2958.2004.04200.x
88. Meibom KL, Li XB, Nielsen AT, Wu C-Y, Roseman S, Schoolnik GK. The *Vibrio cholerae* chitin utilization program. Proc Natl Acad Sci U S A. 2004;101: 2524 LP – 2529. doi:10.1073/pnas.0308707101
89. Weiser JN, Ferreira DM, Paton JC. *Streptococcus pneumoniae*: Transmission,

colonization and invasion. Nat Rev Microbiol. 2018;16: 355–367. doi:10.1038/s41579-018-0001-8

90. Rigali S, Titgemeyer F, Barends S, Mulder S, Thomae AW, Hopwood DA, et al. Feast or famine: the global regulator DasR links nutrient stress to antibiotic production by *Streptomyces*. EMBO Rep. 2008;9: 670–675. doi:https://doi.org/10.1038/embor.2008.83

91. Veening J-W, Blokesch M. Interbacterial predation as a strategy for DNA acquisition in naturally competent bacteria. Nat Rev Microbiol. 2017;15: 621–629. doi:10.1038/nrmicro.2017.66

92. Iyer LM, Aravind L. The catalytic domains of thiamine triphosphatase and CyaB-like adenylyl cyclase define a novel superfamily of domains that bind organic phosphates. BMC Genomics. 2002;3: 33. doi:10.1186/1471-2164-3-33

93. Vogt MS, Ngouoko Nguepbeu RR, Mohr MKF, Albers S-V, Essen L-O, Banerjee A. The archaeal triphosphate tunnel metalloenzyme *Sa*TTM defines structural determinants for the diverse activities in the CYTH protein family. J Biol Chem. 2021;297. doi:10.1016/j.jbc.2021.100820

94. Mamou G, Malli Mohan GB, Rouvinski A, Rosenberg A, Ben-Yehuda S. Early Developmental Program Shapes Colony Morphology in Bacteria. Cell Rep. 2016;14: 1850–1857. doi:https://doi.org/10.1016/j.celrep.2016.01.071

95. Croucher NJ, Finkelstein JA, Pelton SI, Parkhill J, Bentley SD, Lipsitch M, et al. Population genomic datasets describing the post-vaccine evolutionary epidemiology of *Streptococcus pneumoniae*. Sci Data. 2015;2: 150058. doi:10.1038/sdata.2015.58

96. Macfadyen LP, Chen D, Vo HC, Liao D, Sinotte R, Redfield RJ. Competence development by *Haemophilus influenzae* is regulated by the availability of nucleic acid precursors. Mol Microbiol. 2001;40: 700–707. doi:https://doi.org/10.1046/j.1365-2958.2001.02419.x

97. Stewart GJ, Carlson CA. The Biology of Natural Transformation. Annu Rev Microbiol. 1986;40: 211–231. doi:10.1146/annurev.mi.40.100186.001235

98. Redfield RJ. Genes for breakfast: The have-your-cake and-eat-it-too of bacterial transformation. J Hered. 1993;84: 400–404. doi:10.1093/oxfordjournals.jhered.a111361

99. Solomon JM, Grossman AD. Who's competent and when: regulation of natural genetic competence in bacteria. Trends Genet. 1996;12: 150–155. doi:https://doi.org/10.1016/0168-9525(96)10014-7

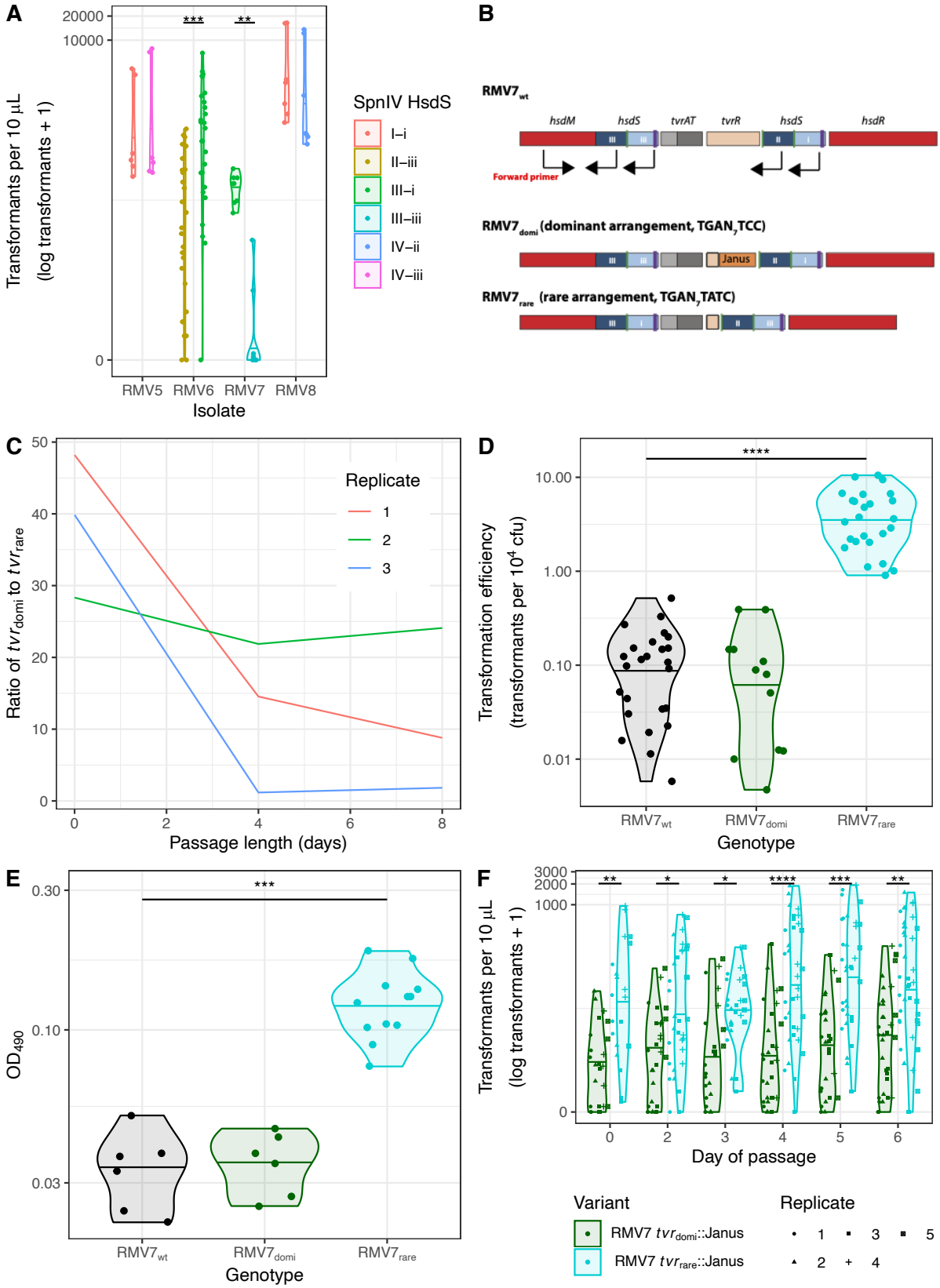
100. Seto H, Tomasz A. Calcium-requiring step in the uptake of deoxyribonucleic acid molecules through the surface of competent pneumococci. J Bacteriol. 1976;126: 1113–1118. doi:10.1128/jb.126.3.1113-1118.1976

101. Gamer J, Multhaup G, Tomoyasu T, McCarty JS, Rüdiger S, Schönfeld HJ, et al. A cycle of binding and release of the DnaK, DnaJ and GrpE chaperones regulates activity of the *Escherichia coli* heat shock transcription factor sigma32. EMBO J. 1996;15: 607–617. doi:https://doi.org/10.1002/j.1460-2075.1996.tb00393.x
102. Selvaraj S, Liu K, Robinson AM, Epstein VA, Conley DB, Kern RC, et al. *In Vivo* Determination of Mouse Olfactory Mucus Cation Concentrations in Normal and Inflammatory States. PLoS One. 2012;7: e39600. Available: https://doi.org/10.1371/journal.pone.0039600
103. Rossol M, Pierer M, Raulien N, Quandt D, Meusch U, Rothe K, et al. Extracellular Ca^{2+} is a danger signal activating the NLRP3 inflammasome through G protein-coupled calcium sensing receptors. Nat Commun. 2012;3: 1329. doi:10.1038/ncomms2339
104. Maier B. Competence and Transformation in *Bacillus subtilis*. Current Issues in Molecular Biology. 2020. doi:10.21775/cimb.037.057
105. Haycocks JRJ, Warren GZL, Walker LM, Chlebek JL, Dalia TN, Dalia AB, et al. The quorum sensing transcription factor AphA directly regulates natural competence in *Vibrio cholerae*. PLOS Genet. 2019;15: e1008362. Available: https://doi.org/10.1371/journal.pgen.1008362
106. Seitz P, Blokesch M. Cues and regulatory pathways involved in natural competence and transformation in pathogenic and environmental Gram-negative bacteria. FEMS Microbiol Rev. 2013;37: 336–363. doi:10.1111/j.1574-6976.2012.00353.x
107. Bosch AATM, Biesbroek G, Trzcinski K, Sanders EAM, Bogaert D. Viral and Bacterial Interactions in the Upper Respiratory Tract. PLOS Pathog. 2013;9: e1003057. Available: https://doi.org/10.1371/journal.ppat.1003057
108. Jeffery CJ. Protein moonlighting: what is it, and why is it important? Philos Trans R Soc B Biol Sci. 2018;373: 20160523. doi:10.1098/rstb.2016.0523
109. Maughan H, Redfield RJ. Extensive variation in natural competence in *Haemophilus influenzae*. Evolution (N Y). 2009;63: 1852–1866. doi:https://doi.org/10.1111/j.1558-5646.2009.00658.x
110. Apagyi KJ, Fraser C, Croucher NJ. Transformation asymmetry and the evolution of the bacterial accessory genome. Mol Biol Evol. 2018;35: 575–581. doi:10.1093/gbe/evs009
111. Elzhov T V, Mullen KM, Spiess A-N, Bolker B. minpack.lm: R Interface to the Levenberg-Marquardt nonlinear least-squares algorithm found in MINPACK, plus support for bounds. 2016. Available: https://cran.r-project.org/package=minpack.lm
112. Fox J, Weisberg S. An {R} Companion to Applied Regression. Third. Thousand Oaks {CA}: Sage; 2019. Available:

https://socialsciences.mcmaster.ca/jfox/Books/Companion/
113. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics. 2013;29: 2933–2935. doi:10.1093/bioinformatics/btt509
114. Kalvari I, Argasinska J, Quinones-Olvera N, Nawrocki EP, Rivas E, Eddy SR, et al. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. Nucleic Acids Res. 2018;46: D335–D342. doi:10.1093/nar/gkx1038
115. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016;34: 525–527. doi:10.1038/nbt.3519
116. Pimentel H, Bray NL, Puente S, Melsted P, Pachter L. Differential analysis of RNA-seq incorporating quantification uncertainty. Nat Methods. 2017;14: 687–690. doi:10.1038/nmeth.4324
117. Guy L, Roat Kultima J, Andersson SGE. genoPlotR: comparative gene and genome visualization in R. Bioinformatics. 2010;26: 2334–2335. doi:10.1093/bioinformatics/btq413
118. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize implements and enhances circular visualization in R . Bioinformatics. 2014;30: 2811–2812. doi:10.1093/bioinformatics/btu393
119. Wilke C. cowplot: Streamlined Plot Theme and Plot Annotations for “ggplot2.” 2020.
120. Kassambara A. ggpubr: “ggplot2” Based Publication Ready Plots. 2020. Available: https://cran.r-project.org/package=ggpubr
121. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the Tidyverse. J Open Source Softw. 2019;4: 1686. doi:10.21105/joss.01686
122. Sobetzko P, Jelonek L, Strickert M, Han W, Goesmann A, Waldminghaus T. DistAMo: A Web-Based Tool to Characterize DNA-Motif Distribution on Bacterial Chromosomes. Front Microbiol. 2016;7: 283. doi:10.3389/fmicb.2016.00283
123. Croucher NJ, Campo JJ, Le TQ, Liang X, Bentley SD, Hanage WP, et al. Diverse evolutionary patterns of pneumococcal antigens identified by pangenome-wide immunological screening. Proc Natl Acad Sci U S A. 2017;114: E357–E366. doi:10.1073/pnas.1613937114
124. Letunic I, Khedkar S, Bork P. SMART: recent updates, new developments and status in 2020. Nucleic Acids Res. 2021;49: D458–D460. doi:10.1093/nar/gkaa937
125. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol. 2013;30: 772–780. doi:10.1093/molbev/mst010
126. Price MN, Dehal PS, Arkin AP. FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS One. 2010;5: e9490.
127. Schoch CL, Ciufo S, Domrachev M, Hotton CL, Kannan S, Khovanskaya R, et al.

NCBI Taxonomy: a comprehensive update on curation, resources and tools. Database (Oxford). 2020;2020. doi:10.1093/database/baaa062

1 **Figures**



2

3 **Figure 1** Variation in transformability between locked *tvr* variants. (A) Violin plot showing the

4 transformation efficiency of four pairs of *tvr* locus variants constructed from isolates RMV5,

5 RMV6, RMV7 and RMV8. Each individual point represents an independent transformation

experiment. The horizontal line within each violin shows the median for each genotype. (B) Schematic of the *tvr* loci from $RMV7_{wt}$, $RMV7_{domi}$ and $RMV7_{rare}$ to show the genes encoding the methylase (*hsdM*), endonuclease (*hsdR*), regulatory system (*tvrAT*) and recombinase (*tvrR*). The variants differ in their active *hsdS* genes, upstream of *tvrATR*. The $RMV7_{wt}$ and $RMV7_{domi}$ HsdS proteins comprise the TRD combination III-iii (recognising motif $TGAN_7TCC$), whereas that of $RMV7_{rare}$ contains the TRDs III-I (recognising motif $TGAN_7TATC$). The black arrows indicate the binding sites of a forward primer, in *hsdM*, and reverse primers, in *hsdS* fragments. (C) Line graph showing the ratio of $RMV7_{domi}$ to $RMV7_{rare}$ loci in eight-day passages of $RMV7_{wt}$. Although conversion to $RMV7_{rare}$ was detected, the $RMV7_{domi}$ variant was more common in each of the three replicates at every measured timepoint. (D) Violin plot showing the higher transformation efficiency of RMV_{rare} relative to $RMV7_{domi}$ or $RMV7_{wt}$. Each individual point represents an independent transformation experiment in which the number of transformants, and number of overall colony-forming units (cfu), was calculated. The horizontal line within each violin shows the median for each genotype. Both mutants were compared with $RMV7_{wt}$; the horizontal bracket shows a significant difference, as calculated from a two-tailed Wilcoxon rank sum test. (E) Violin plot showing the adhesion of variants to an abiotic surface, as quantified by OD_{490} after crystal violet staining of cells resisting washing from a microplate well. The $RMV7_{rare}$ variant showed a significantly greater adhesion after 16 hours incubation at 35 °C, as assessed by a Wilcoxon rank sum test. (F) Violin plots showing the transformation efficiency of knock-in mutants during a passage experiment. The *tvr* loci of RMV7 tvr_{domi}::Janus and RMV7 tvr_{rare}::Janus (both containing a *tvrR* gene interrupted by a Janus cassette) were each introduced into an $RMV7_{wt}$ background. This isogenic pair were separately passaged in liquid cultures over six days in five independent replicates. The number of transformants observed from three transformation assays, conducted each day for both variants, is shown by the individual points' shapes and colours. The violin plots summarise the median and distribution of these values. The brackets indicate the statistical significance of the comparison between variants from the same day of the passage, as calculated using a Wilcoxon rank sum test. Across all panels, significance is coded as: $p < 0.05$, *; $p < 0.01$, **; $p < 10^{-3}$, ***; $p < 10^{-4}$, ****. All p values were subject to a Holm-Bonferroni correction within each panel.

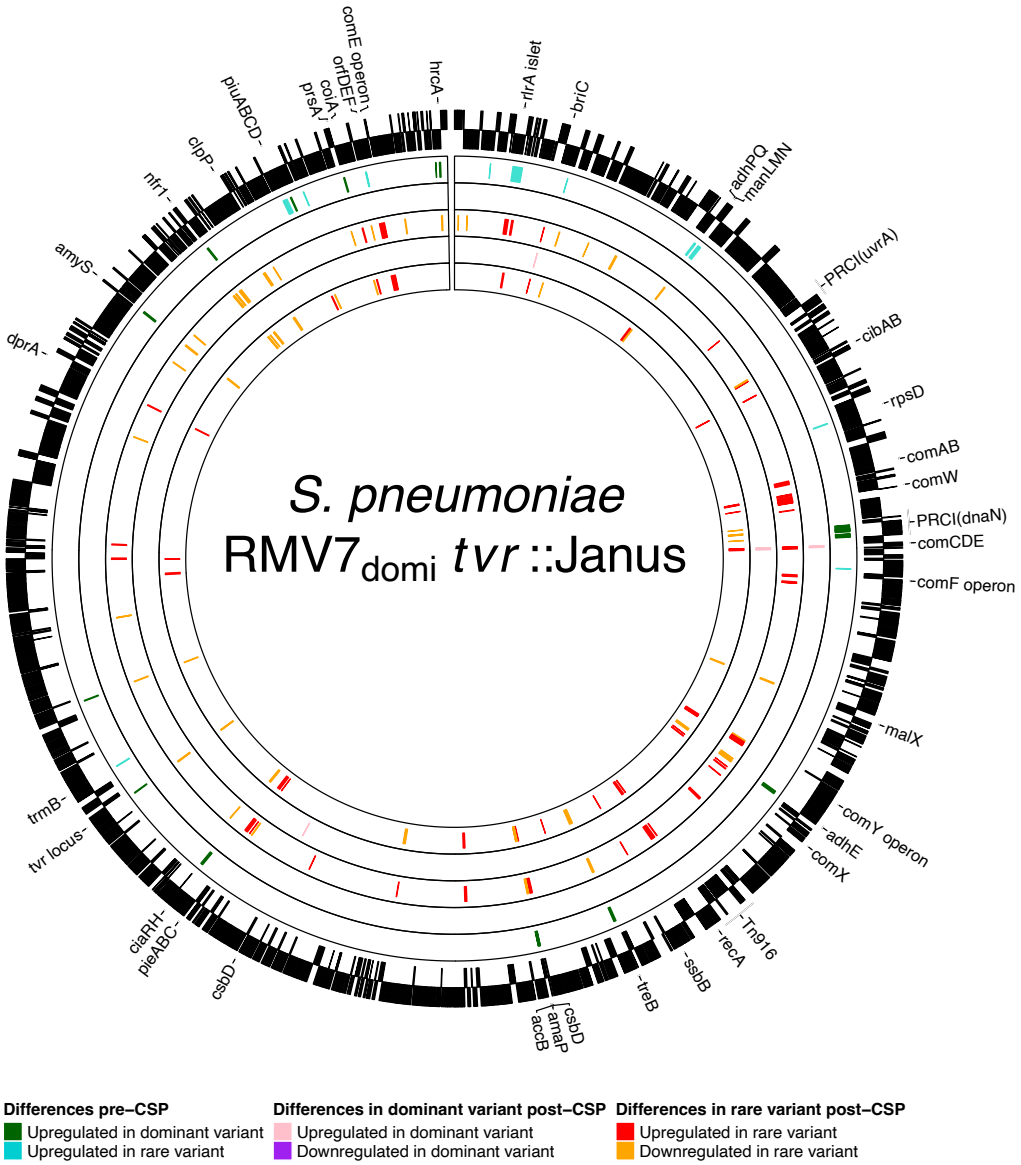


Figure 2 Chromosomal distribution of genes exhibiting significant differences in transcription between RNA-seq samples. The outer ring shows the annotation of $RMV7_{domi}$. Protein coding sequences are represented as black boxes, with the vertical positioning within the ring indicating the strand of the genome on which they are encoded. The next ring inwards shows significant pre-CSP differences in transcription: green genes were more highly expressed in RMV_{domi}, and blue genes were more highly expressed in RMV_{rare}. The next ring inwards show significant changes in gene expression 10 minutes post-CSP in $RMV7_{domi}$: pink genes were upregulated, and purple genes were downregulated. The third ring inwards shows significant changes in gene expression 10 minutes post-CSP in $RMV7_{rare}$: red genes were upregulated, and orange genes were downregulated. The two inner rings repeat this representation for changes in gene expression 20 minutes post-CSP.

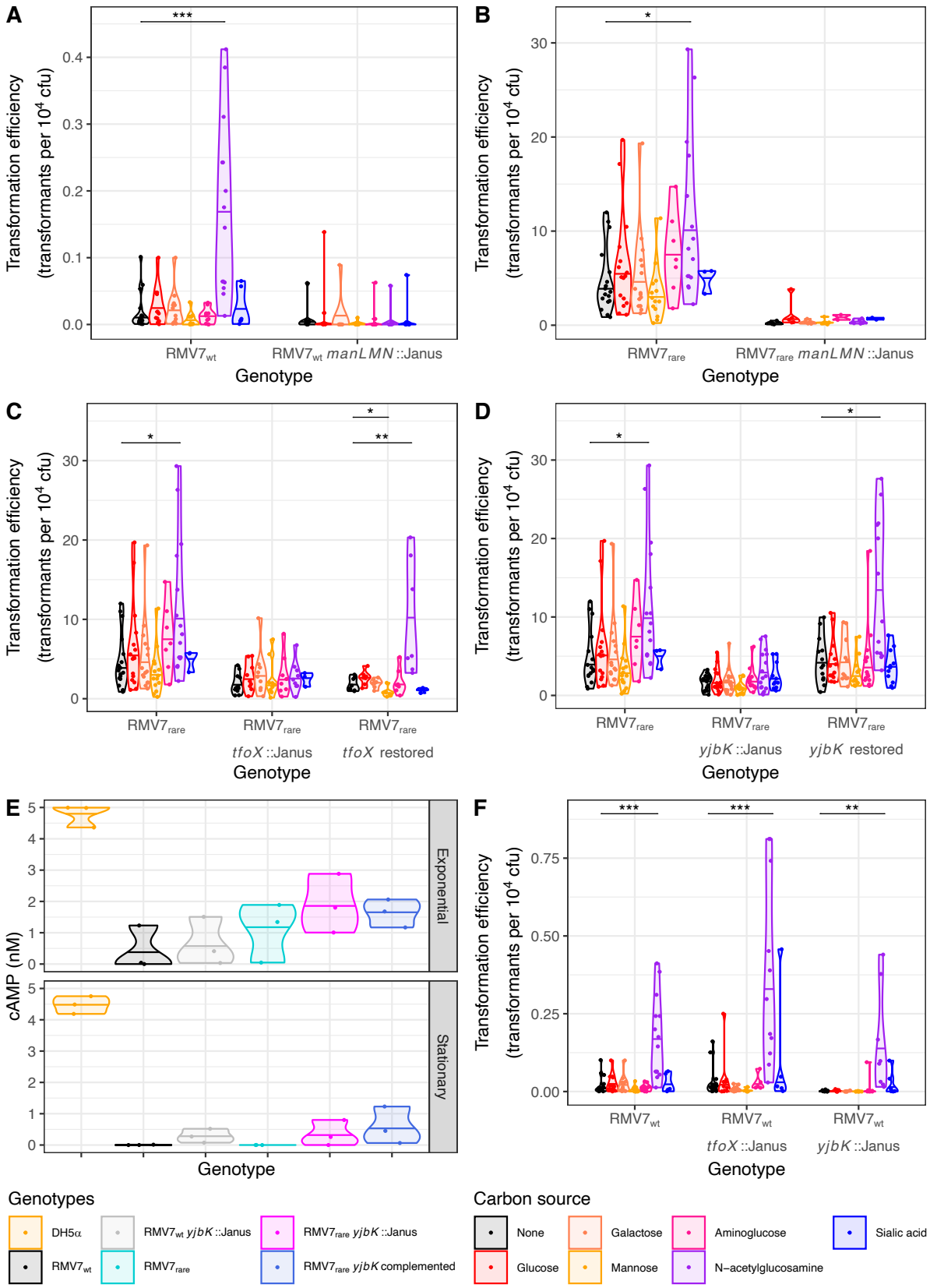


Figure 3 The dependence of transformation efficiency on import of carbon sources. (A) Violin plots showing the transformation efficiency of $RMV7_{wt}$ relative to a mutant in which *manLMN* had been disrupted by a Janus cassette. Each genotype was transformed in unsupplemented media, and in the presence of one of six carbon sources, as indicated by

the plot colour (see key on bottom right). Each point represents an independent experiment, and the horizontal line within the violin plots show the median for each combination of recipient cell genotype and carbon source. For each genotype, Wilcoxon rank-sum tests were used to test for evidence of changes in transformation efficiency caused by each carbon source. Significant differences are indicated by the black brackets at the top of the panel. (B) Violin plots showing the transformation efficiency of $RMV7_{rare}$ relative to a mutant in which *manLMN* had been disrupted by a Janus cassette. Data are displayed as in panel A. (C) Violin plots showing the transformation efficiency of $RMV7_{rare}$ relative to a mutant in which *tfoX* had been disrupted by a Janus cassette, and a third genotype in which *tfoX* had been restored. Data are displayed as in panel A; the data for $RMV7_{rare}$ are replicated from panel B. (D) Violin plots showing the transformation efficiency of $RMV7_{rare}$ relative to a mutant in which *yjbK* had been disrupted by a Janus cassette, and a third genotype in which *yjbK* had been restored. Data are displayed as in panel A; the data for $RMV7_{rare}$ are replicated from panel B. (E) Violin plots showing the concentration of 3',5'-cAMP in samples taken from exponential and stationary phase cultures of *E. coli* DH5α, and *S. pneumoniae* $RMV7_{wt}$ and $RMV7_{rare}$ genotypes differing in whether *yjbK* was intact or not. The genotypes are indicated by the colours (see key on bottom left). (F) Violin plots showing the transformation efficiency of $RMV7_{wt}$ relative to mutants in which either *tfoX* or *yjbK* had been disrupted by a Janus cassette. Data are displayed as in panel A. Across all panels, significance is coded as: $p < 0.05$, *; $p < 0.01$, **; $p < 10^{-3}$, ***; $p < 10^{-4}$, ****. All p values were subject to a Holm-Bonferroni correction within each panel.

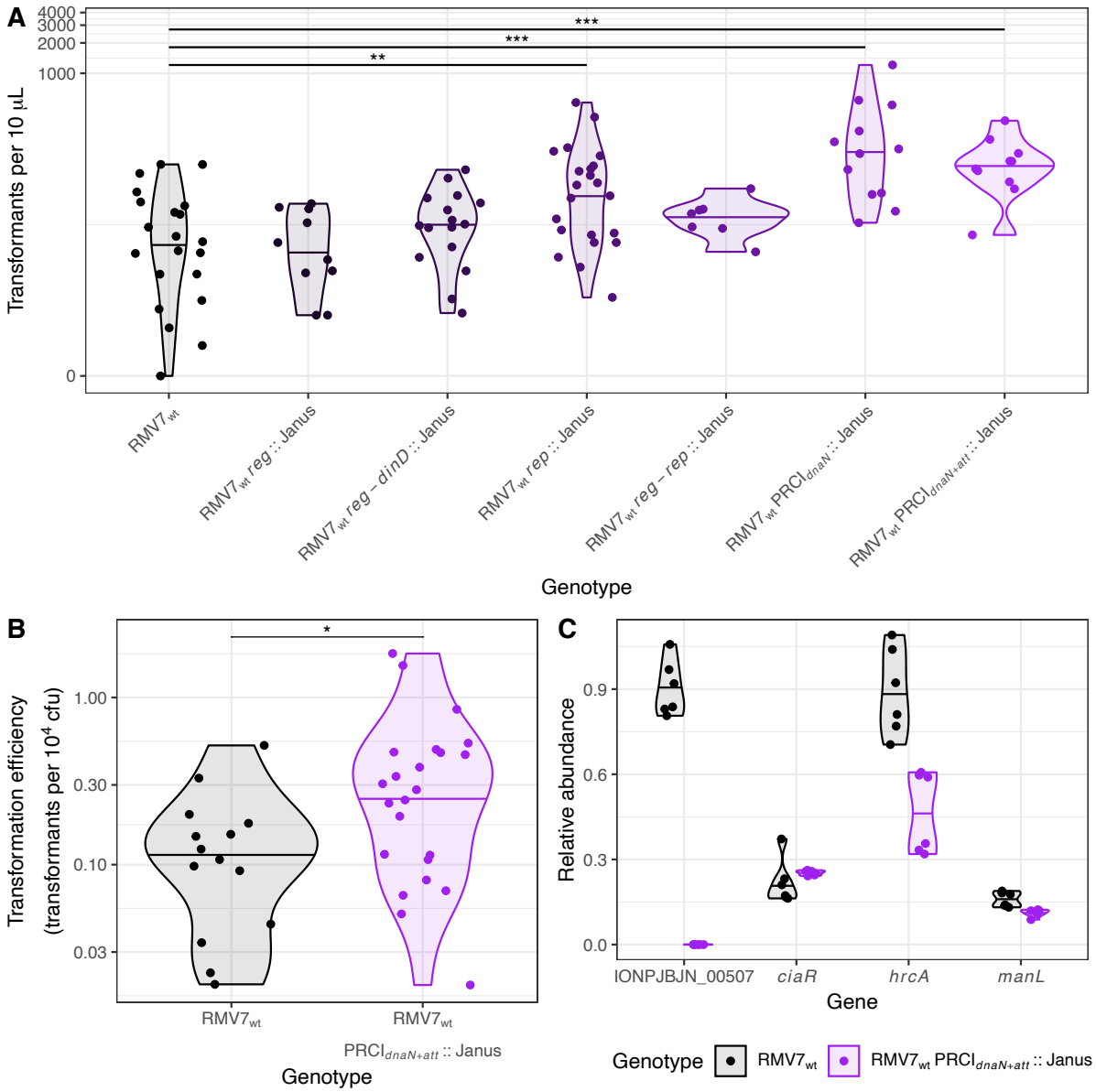


Figure 4. Effect of removing PRCI$_{dnaN}$ on RMV7$_{wt}$. (A) Violin plots showing the number of transformants observed in assays of RMV7$_{wt}$ mutants in which different parts of PRCI$_{dnaN}$ were replaced with a Janus cassette. The genotypes are arranged, and coloured, according to how much of the element was replaced by the cassette. Each point is an independent transformation assay. The violin plot summarises the result for each mutant, with a horizontal line indicating the median. Each mutant was compared against the parental RMV7$_{wt}$ genotype using a Wilcoxon rank sum test. Black brackets at the top of the plot indicate significant differences in the number of observed transformants. (B) Violin plot quantifying the effect of PRCI$_{dnaN}$ on transformation efficiency. This comparison of RMV7$_{wt}$ with a mutant in which the PRCI and its *att* site were removed was independent of the experiments presented in panel A, and more accurately quantified transformation efficiency. The significant difference, calculated from a two-tailed Wilcoxon rank sum test, is indicated by the bracket at the top of the panel. (C) Violin plots showing gene expression, as

quantified by qRT-PCR, in the genotypes assayed in panel B. IONPJBJN_00507 is a coding sequence within the PRCI that is absent from $RMV7_{wt}$ $PRCI_{dnaN+att}$::Janus. The six points for each gene correspond to three technical replicate assays on each of two biological replicates. The horizontal line on the violin plot shows the median relative abundance for each gene in each genotype. Across all panels, significance is coded as: $p < 0.05$, *; $p < 0.01$, **; $p < 10^{-3}$, ***; $p < 10^{-4}$, ****. All p values were subject to a Holm-Bonferroni correction within each panel.

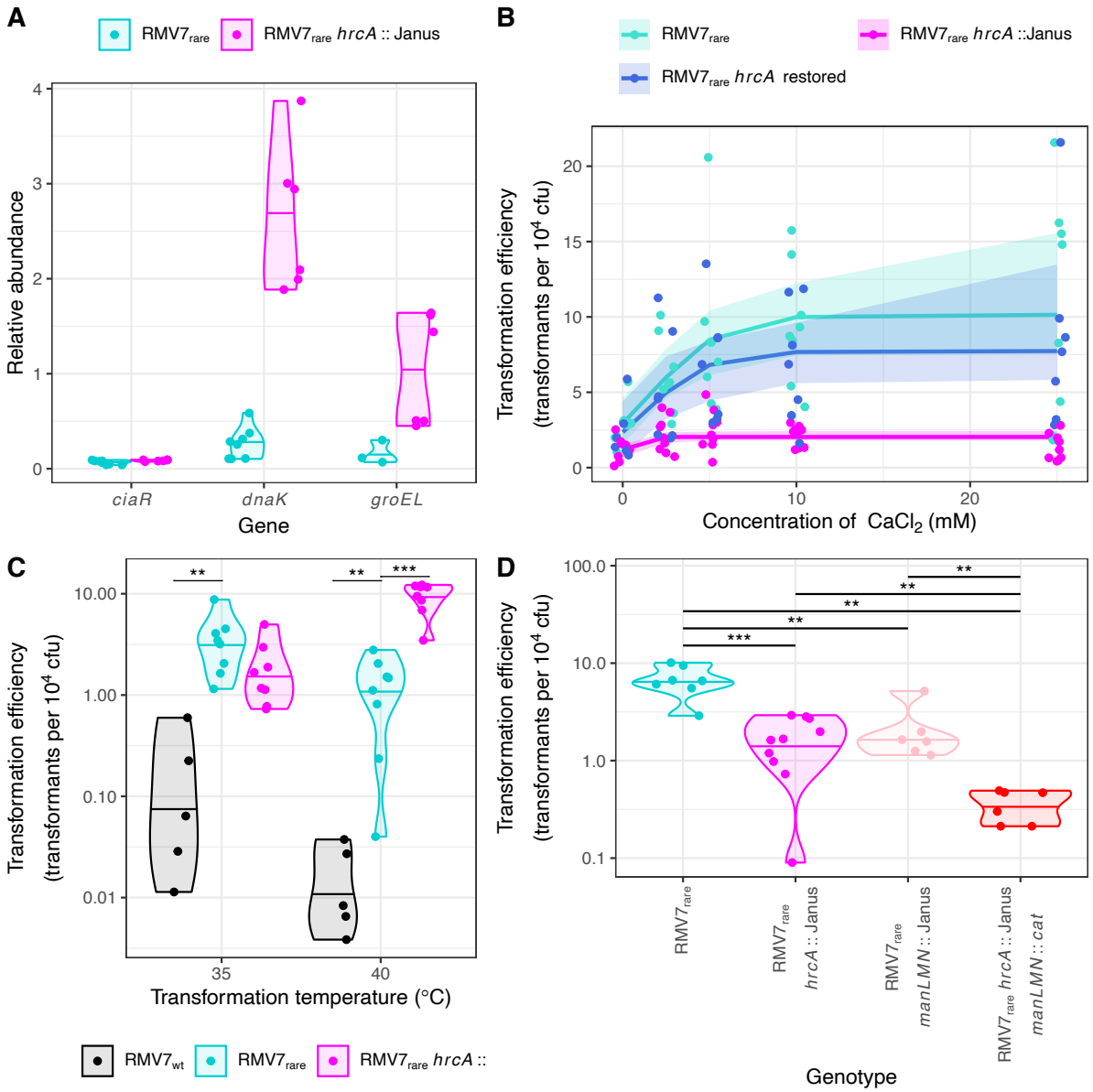


Figure 5 The regulation of transformation by HrcA in $RMV7_{rare}$. (A) Violin plots showing gene expression, as quantified by qRT-PCR, in $RMV7_{rare}$ and a mutant derivative in which the *hrcA* chaperone regulator gene was disrupted by a Janus cassette. The six points for each gene correspond to three technical replicate assays of each of two biological replicates. The horizontal line on the violin plot shows the median relative abundance for each gene in each genotype. The absence of any change in *ciaR* expression demonstrated the effect on transformation was independent of CiaRH. (B) Scatterplot showing the effect of $CaCl_2$ on transformation efficiency in $RMV7_{rare}$ and mutants in which *hrcA* had been disrupted, and then restored. Each point represents an independent transformation assay of one genotype at the indicated $CaCl_2$ concentration. The best-fitting dose response logistic models are shown, with the shaded areas corresponding to the 95% confidence intervals. (C) Violin plots showing the transformation efficiency of $RMV7_{wt}$, $RMV7_{rare}$ and $RMV7_{rare}$ *hrcA*::Janus

during normal growth (35 °C) or following a 40 °C heat shock. Each point corresponds to an independent transformation experiment, and the violin plots have a horizontal line indicating the median transformation efficiency of each mutant at each temperature. (D) Combined effects of chaperone and carbon source regulation on transformation efficiency. Results are displayed as in panel C. Wilcoxon rank-sum tests were conducted between all pairs of genotypes. These found both the single mutants, lacking *manLMN* or *hrcA*, were significantly less transformable than the parental genotype. Furthermore, the double mutant was less transformable than either single mutant. Across all panels, significance is coded as: $p < 0.05$, *; $p < 0.01$, **; $p < 10^{-3}$, ***; $p < 10^{-4}$, ****. All p values were subject to a Holm-Bonferroni correction within each panel.

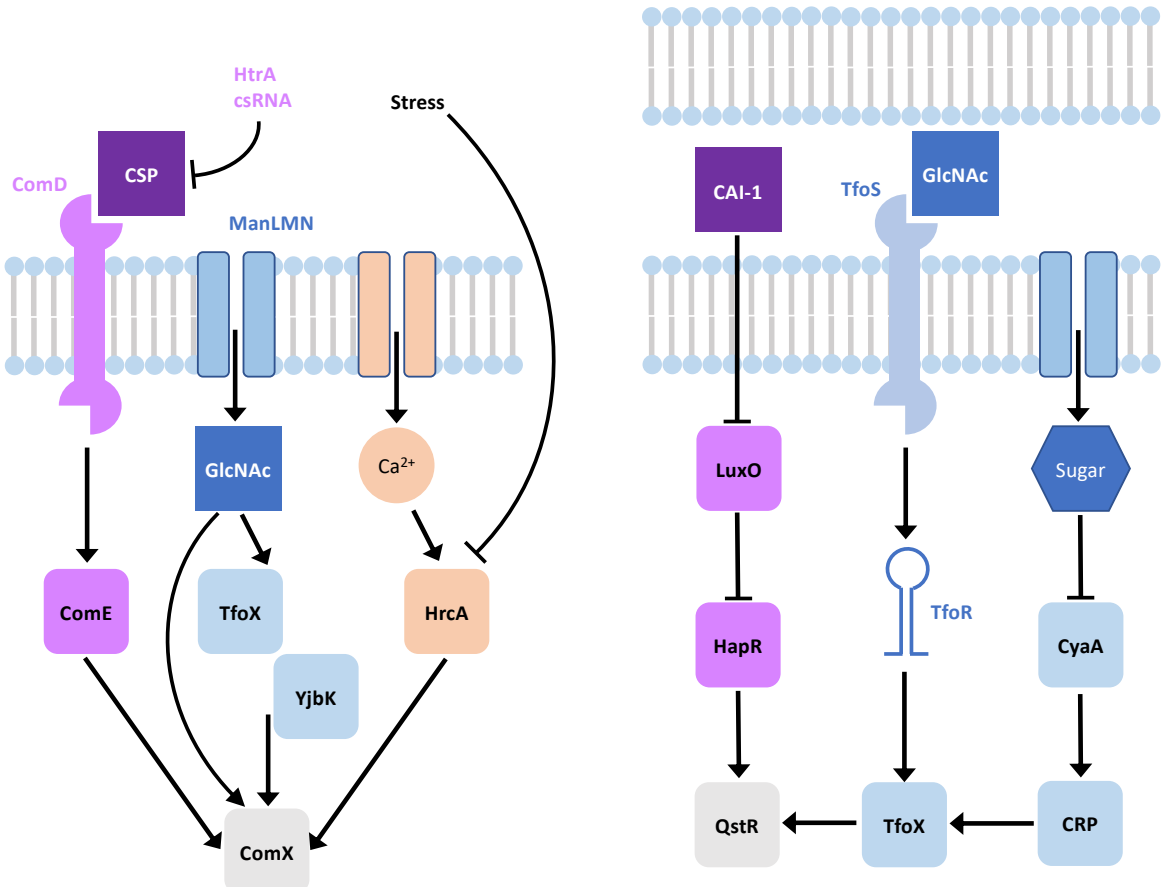


Figure 6 Comparison of the regulation of competence in *S. pneumoniae* (left), from this work, and *V. cholerae* (right), summarized from [71]. In each, competence is regulated by a quorum-sensing system: CSP in *S. pneumoniae*, and the cholera autoinducer 1 (CAI-1) in *V. cholerae*. CAI-1 operates through inhibiting LuxO, thereby activating the HapR protein, which indicates a high-cell density environment. HapR activates competence through the quorum-sensing and TfoX-dependent regulator (QstR). This regulator also senses the activation of TfoX in response to GlcNAc being sensed by the transmembrane regulator TfoS, via the TfoR small RNA. TfoX activity is regulated by the Catabolite Regulatory Protein (CRP), which is activated by 3',5'-cAMP, generated by the CyaA adenylate cyclase under carbon source starvation conditions. Hence there are parallels with the TfoX orthologue, and adenylate cyclase-like protein YjbK, responding to GlcNAc in *S. pneumoniae* RMV7. GlcNAc also appears to promote competence through a TfoX/YjbK-independent route, based on the behaviour of $RMV7_{wt}$. There are no such parallels for the regulation of competence by HrcA, the binding of which to CIRCE sequences is promoted by Ca^{2+}, thereby activating competence through an unknown mechanism. Note in *S. pneumoniae*, these signals are shown to converge on the competence-specific sigma factor ComX for simplicity, but the mechanisms and targets of the signalling pathways are unknown.