Abstract *Streptococcus pneumoniae* can cause disease in various human tissues and organs, including the ear, the brain, the blood and the lung, and thus in highly diverse and changing environments. It is challenging to study how pneumococci control virulence factor expression because cues of natural environments and the presence of an immune system are difficult to simulate in the lab. Here, we apply synthetic biology methods to reverse engineer gene expression control in *S. pneumoniae*. A selection platform is described that allows for straightforward identification of regulatory elements out of combinatorial libraries. Furthermore, we demonstrate TetR- and LacI-regulated promoters that show dynamic ranges of four orders of magnitude. On the basis of these promoters, regulatory networks of higher complexity are assembled, such as inverters and a logic AND gate. Finally, we present single-copy genomic integrated toggle switches that give rise to bimodal population distributions. The tools described here pave the way for *in vivo* investigations of the importance of gene expression control on the pathogenicity of *S. pneumoniae* and can be used to mimic expression patterns of essential virulence factors.
Introduction

Human pathogens and commensals reside in highly dynamic environments where they are required to interact with host cells and the immune system, and with niche competitors. *Streptococcus pneumoniae* (pneumococcus) is a prominent example of a colonizer of such complex habitats. The Gram-positive bacterium is generally found in a commensal state in the human nasopharynx; however, pneumococci can also cause disease, such as otitis media, meningitis, sepsis and pneumonia, and they are responsible for more than one million deaths per year \(^1,2\). To date, it remains unclear why *S. pneumoniae* represents such an important health threat, and how genes that enable pneumococci to fight or escape the immune system are regulated \(^3\).

In order to be a successful colonizer (and subsequent pathogen), *S. pneumoniae* needs to tightly control its gene expression in response to both internal signals, such as the cell cycle, and external signals, such as nutrient availability, the inhibition by immune system components or the presence of competitors. Although hundreds of studies have been dedicated to unraveling the regulons of key regulators \(^4–8\), it is still unclear how pneumococci manage to accurately control virulence factor gene expression under changing conditions. It is particularly challenging to investigate these phenomena because dynamic *in vivo* environments are difficult to simulate in the lab. Furthermore, gene expression strategies often involve population bifurcations and thus take place at the single-cell level, giving rise to various subpopulations of specialized phenotypes \(^9,10\).

One strategy to investigate complex gene expression patterns relies in the approach of understanding by engineering, and thus in studying artificial regulatory networks consisting of orthogonal components that mimic natural networks \(^11\). Most of this work in bacteria has been performed in model organisms, such as *Escherichia coli* and *Bacillus subtilis* \(^12–17\); for studies of human pathogens, these tools mostly still need to be developed. We recently introduced a platform for synthetic biology approaches in *S. pneumoniae* \(^18\). The integration plasmid pPEP enables the assembly of genetic elements via the BglBrick cloning system \(^19\). On the basis of this platform, gene expression regulation in general, and more complex gene expression patterns, such as the ones observed for virulence factors, can be investigated. A bimodal population distribution, for example, which is a common feature of virulence gene expression \(^9,20\), could be studied on the basis of a synthetic bistable switch. Artificial regulatory networks would allow for examining the importance of complex patterns of virulence factor expression *in vivo*, such as in a mouse infection model, and thus in the presence of a functional immune system.

The most fundamental elements for the control of gene expression are promoters that regulate transcription. The basic architecture of prokaryotic promoters has been the focus of
SYNTHETIC BIOLOGY TOOLS TO CONTROL GENE EXPRESSION IN *STREPTOCOCCUS PNEUMONIAE*

numerous studies. Processes such as initial promoter recognition by transcription factors and RNA polymerase (RNAP), transcription initiation, promoter escape and transcription elongation are well-characterized\(^{21,22}\), and have even been addressed recently in *S. pneumoniae*\(^{23}\). Nevertheless, to date, transcription is not sufficiently understood to allow for accurate computational simulations, and it can therefore also not be engineered in a predictable manner\(^{24}\). The identification of a suitable promoter for the expression of a specific gene in a given genetic background, or the construction of an artificial regulatory network of determined function will remain a trial and error process as long as knowledge limitations preclude accurate simulations. A good strategy to shortcut such trial and error processes can be found, for example, in selection assays of genetic libraries\(^{11,25}\).

Here, we apply synthetic biology approaches to engineer gene expression in *S. pneumoniae*. A selection platform was created that enables the identification of genetic elements for specific gene expression patterns out of comprehensive libraries. Orthogonal transcription factors were introduced and functionalized in *S. pneumoniae*. On the basis of these regulators, complex gene expression networks were assembled and characterized, such as inverters, a logic AND gate and toggle switches. These tools can be applied for *in vivo* investigations of the contribution of gene expression control on the pathogenicity of *S. pneumoniae*.

**Results**

**Selection and counterselection systems for *S. pneumoniae***

Synthetic gene regulatory networks are difficult to construct to date in *S. pneumoniae* or even in model organisms because of a lack of well-characterized individual parts, and because of limited knowledge in predicting interference between assembled parts\(^{24,26,27}\). To bypass these limitations, we aimed at developing an alternative strategy: the identification of individual parts and assembled gene regulatory networks based on the selection and counterselection of genetic libraries. With such a system, parts of desired functionality, such as strong promoters, could be identified via selection. For example, promoter libraries can be cloned in front of an antibiotic resistance marker followed by treatment with antibiotics\(^{28}\). Conversely, weak promoters can be identified via counterselection\(^{29}\), and more complex transcription patterns can potentially be found via sequential selection and counterselection.

An ideal positive selector would have a large dynamic range related to the inducer concentration, similar growth behavior of cells in liquid culture and on agar plates, and limited spontaneous resistance emergence. In order to create a positive selection platform
for *S. pneumoniae*, we first tested and cloned a set of five antibiotic resistance markers under the control of the Zn\(^{2+}\)-inducible promoter PZ1 (Fig. 1a). As expected, cells only grew at high concentrations of antibiotics when PZ1 was induced with high concentrations of Zn\(^{2+}\) (Fig. 1b). Within our applied resistance marker induction series, pneumococcal populations responded to concentrations of chloramphenicol and tetracycline within one order of magnitude, kanamycin spanned two orders of magnitude, and both erythromycin and trimethoprim showed a response range of more than three orders of magnitude (Fig. 1b). Since trimethoprim selection worked less well on plates, and is furthermore prone for the emergence of spontaneously resistant cells (by SNPs in *dhfR*), we chose erythromycin and its marker *erm* as positive selector.

To date, and to our knowledge, there is only one counterselectable system described for *S. pneumoniae* called Janus\(^30\). Janus cloning relies on a streptomycin resistant strain that becomes susceptible when the wild type allele of the ribosomal gene *rpsL* is expressed. Janus was furthermore extended by adding the *Bacillus subtilis* gene *sacB* (Sweet Janus)\(^31\) which confers sucrose sensitivity (via an unknown mechanism)\(^32\); however, counterselection with *sacB* was not shown to work independently of the Janus context in *S. pneumoniae*\(^31\). The disadvantage of Janus cloning is the requirement of a genetic background that carries a mutated *rpsL* allele. The *pheS* counterselection system of *E. coli* does not rely on a mutated genetic background\(^33\). Instead, the *pheS*\(^{A294G}\) *E. coli* allele, encoding a phenylalanine-tRNA ligase, is dominant negative over wild type *pheS*. Expression of *pheS*\(^{A294G}\) allows for the incorporation of a toxic analog of phenylalanine, called para-chlorophenylalanine, into proteins which in turn causes growth retardation. We introduced this system into *S. pneumoniae* by cloning the pneumococcal equivalent of *E. coli* *pheS*\(^{A294G}\), *S. pneumoniae* *pheS*\(^{A315G}\), under control of PZ1 (Fig. 1a). Indeed, only in the presence of para-chlorophenylalanine the expression of *pheS*\(^{A315G}\) led to reduced growth in liquid culture (Fig. 1d). For reasons currently unknown the expression of *pheS*\(^{A315G}\) did not retard bacterial growth on agar plates containing 2 mg ml\(^{-1}\) of para-chlorophenylalanine.

After the identification of a suitable selection system (*erm*/erythromycin), and a counterselection system for liquid culture (*pheS*/para-chlorophenylalanine), we combined the two markers together with *luc* (firefly luciferase), allowing for the analysis of gene expression at the population level, and together with *gfp* (green fluorescent protein), allowing for single-cell analysis. BglBrick cloning was used to assemble genes in a consecutive manner into the vector pPEP (see Methods). The resulting selection platform, named pPEP7, was first tested by placing it under the control of the Zn\(^{2+}\)-inducible promoter PZ1. To do so, plasmid pPEP74 (harboring PZ1-*pheS-erm-luc-gfp*) was transformed into *S. pneumoniae* D39\(^34\) resulting in strain D-PEP74. As shown in Figure 1c,
SYNTHETIC BIOLOGY TOOLS TO CONTROL GENE EXPRESSION IN *STREPTOCOCCUS PNEUMONIAE*

When strain D-PEP74 was grown in the presence of 0.1 µg ml⁻¹ erythromycin, a clear Zn²⁺-dependent growth profile was observed. In contrast, when D-PEP74 was grown in the presence of 2 mg ml⁻¹ *para*-chlorophenylalanine, the dose–response relationship of Zn²⁺ induction was inverse; the absence of Zn²⁺ allowed for uninhibited growth while high Zn²⁺ levels resulted in growth arrest (Fig. 1d). However, after a lag period of approximately 7 h,
Zn\textsuperscript{2+}-induced cells restarted growth (Fig. 1d). This phenomenon likely originated from mutated subpopulations that overtook the wild type population. Sequencing of four isolates displaying spontaneous loss of \textit{para}-chlorophenylalanine susceptibility revealed that \textit{pheS\textsuperscript{A315G}} was unaltered. However, growth cultures of these strains were found to be less responsive towards Zn\textsuperscript{2+} induction, indicating a mutation event within genes of the regulatory network for Zn\textsuperscript{2+} homeostasis (data not shown).

**Efficient selection of a wide range of constitutive pneumococcal promoters**

We tested the ability of the new selection platform to identify genetic elements of desired function by screening promoter libraries. We reasoned that by changing the selection conditions (either by selecting with erythromycin for high expression, or by selecting with \textit{para}-chlorophenylalanine for weak expression) we could extract promoters of a desired strength out of cloned oligonucleotide libraries. Furthermore, the platform might allow us to test the contribution of individual promoter elements to promoter activity and in that way gain new biological insights into pneumococcal gene expression. DNA sequence variations can influence the promoter strength in various ways. For example, the affinity to RNAP subunits or to transcription factors can be affected, such as the UP element to the α carboxy-terminal domain of RNAP, or the −35 and −10 hexamers to σ\textsuperscript{70}; furthermore, sequence variations can alter the kinetics of open complex formation, transcription initiation and promoter escape\textsuperscript{21}.

We used a strong synthetic promoter (P2\textsuperscript{18}) as a template for four distinct promoter libraries, containing randomized sequences in: (i) the UP element (UP, position −58 to −36 relative to the transcription start site\textsuperscript{35}, resulting in \(4^{23} = 7.0 \times 10^{13}\) potential different promoters); (ii) the core region (CORE, the 17 nucleotides between the −35 and the −10 hexamers, \(4^{17} = 1.7 \times 10^{10}\) potential combinations); (iii) the −35 and −10 hexamers (TATA, only specific sequence variations were allowed\textsuperscript{36}, resulting in \(2^{6} = 64\) potential combinations); (iv) the proximal region (PROX, in this case the 14 nucleotides immediately downstream of the −10 hexamer, \(4^{14} = 2.7 \times 10^{8}\) potential combinations) (Fig. 2a).

Promoter libraries were constructed by polymerase driven extension of two oligonucleotides that overlapped approximately 20 base pairs, with one oligonucleotide containing a randomized section (Fig. 2a, b). Double-stranded promoter constructs were integrated into pPEP7 via BglBrick assembly and sub-cloning in \textit{E. coli} MC1061; 3000 to 4000 colonies were pooled and plasmid DNA was isolated, cut with BglII to eliminate self-ligations (in this case the restriction site disappeared upon promoter integration), and pPEP7 libraries were finally transformed into \textit{S. pneumoniae} D39 (Fig. 2b; see also Methods).
SYNTHETIC BIOLOGY TOOLS TO CONTROL GENE EXPRESSION IN *STREPTOCOCCUS PNEUMONIAE*

D39 libraries underwent four different selection treatments (Fig. 2c). Cells were plated either with spectinomycin (resistance marker of the pPEP backbone) to obtain all possible promoters (Control), or with a high concentration of erythromycin to obtain strong promoters (positive selection). Alternatively, transformed cultures were first treated with *para*-chlorophenylalanine for 4 h (and additional spectinomycin to inhibit non-transformed cells) and subsequently plated either with a low dosage of erythromycin for intermediate
promoters (sequential selection), or with spectinomycin for weak promoters (negative selection). After overnight incubation, cells from 24 individual colonies per selection condition were analyzed by cultivation in microtiter plates. Luminescence from luciferase expression served as readout of the promoter strength.

Transformants originating from the control treatment demonstrated a wide range of promoter strengths, as measured by luminescence (Fig. 3a–d). In case of the UP, CORE and PROX libraries at least half of the control transformants showed high levels of

---

**Figure 3 | Expression strength of selected promoter library strains.** (a–d) Distribution of the promoter strength, measured by luminescence from luciferase expression, of 24 individual strains per selection condition of the UP library (a), the CORE library (b), the TATA library (c), and the PROX library (d); (C), control; (−), negative selection; (−/+), sequential selection; (+), positive selection; ×10^2, expression strength from 1.0×10^2 to 9.9×10^2 RLU OD −1 (relative luminescence units per optical density at 595 nm) (see also Methods). (e) Sequence and expression strength of four promoters of the UP library (in comparison to P2), derived from negative selection (P-B2), sequential selection (P-C7), and positive selection (P-E9 and P-F6); transcription start sites are shown in bold, restriction sites are shown in dark grey, sequence deviations are underlined.
luminescence (1.0×10^5 to 9.9×10^6 RLU OD\(^{-1}\), relative luminescence units per optical density at 595 nm) (Fig. 3a, b, d). In contrast, random sampling of isolates from the TATA library did not yield promoters with similar strength than the template promoter, which carries the consensus TATAAT sequence, confirming the importance of the canonical –35 and –10 sequences for functional promoters (Fig. 3c). Importantly, promoters of a desired strength (strong, intermediate or weak) could be selectively enriched from all four promoter libraries (Fig. 3a–d). For example, when selecting with para-chlorophenylalanine, only transformants showing low levels of luminescence were recovered, while selection in the presence of erythromycin led to the recovery of transformants with high gene expression activity (Fig. 3a–d). To examine whether the libraries were truly diverse in sequence, four promoters of the UP library (one weak, one intermediate, and two strong ones) were sequenced. According to the expectation, all UP element sequences deviated from the P2 template promoter (Fig. 3e). Interestingly, both the weak and the intermediate promoter contained an additional sequence deviation within the –10 hexamer, suggesting that random sequence variations within the UP element did not frequently result in an impairment of the promoter activity of P2. These results indicate that future work on a detailed analysis of recovered promoter sequences might provide new insights into pneumococcal gene expression control.

**TetR- and LacI-regulated promoters**

The above-mentioned results showed that our cloning vector and our selection platform could be successfully applied to identify constitutive promoters of a desired strength. Next, we sought to identify controllable promoters from which transcription can be induced by the exogenous addition of small molecules. To date, there are only few inducible systems available for *S. pneumoniae*, showing different drawbacks for specific applications. These systems are either based on pneumococcal regulators and they are thus not orthogonal\(^{18,37–39}\), or they are regulated by peptides and thus require complex, membrane associated signaling machineries\(^{40,41}\), or they show a limited dynamic range\(^{42}\). We aimed at introducing orthogonal transcription factors into *S. pneumoniae* that are regulated by small diffusible molecules, and that enable a large dynamic range. The most commonly used and best characterized bacterial regulators are the TetR and LacI repressors, which bind in form of dimers to operator sites (*tet*O and *lac*O, respectively) consisting of 19 to 21 base pair-long DNA sequences with dyad symmetry. TetR originates from the tetracycline-resistance operon encoded in Tn10 of *E. coli* and responds to the antibiotic tetracycline\(^{43,44}\) while LacI represses the *lac*-operon and responds to the sugar allolactose\(^{45}\). These compounds can interact with their corresponding repressors and trigger conformational changes that
Figure 4 | ATc- and IPTG-inducible promoters. (a) Sequence of TetR repressed promoters PT and LacI repressed promoters PL, based on the constitutive promoter P2, with TetR operator sequences in light blue and LacI operator sequences in magenta; transcription start sites are shown in bold, restriction sites are shown in dark grey (omitted for most LacI promoters because of space limitations), sequence deviations are underlined. (b, c) Luminescence from luciferase expression, driven by PT promoters (b) and PL promoters (c), without induction and with maximum induction, integrated at the amiF locus together with the strong constitutive promoter PF6 driving regulator expression; ATc, anhydrotetracycline; IPTG, isopropyl β-D-1-thiogalactopyranoside; error bars see Methods. (d, e) Induction series of selected PT promoters with ATc (d) and of selected PL promoters with IPTG (e), measured
dramatically decrease the binding affinity for operator sites. Importantly, non-toxic and non-degradable inducer molecules to these repressors exist, ATc (anhydrotetracycline) in the case of TetR, and IPTG (isopropyl β-D-1-thiogalactopyranoside) in the case of LacI.

First, we codon-optimized tetR and lacI for expression in S. pneumoniae D39 and integrated them, together with an optimal pneumococcal ribosome binding site, upstream of the multiple cloning site of pPEP, under the control of the previously identified strong constitutive promoter PF6 (Fig. 3e). The resulting plasmids, pPEP8T and pPEP8L, respectively, furthermore harbor luc and gfp in the BglBrick cloning site for gene expression analysis (Fig. 4b, c). Oligonucleotide libraries, as were used to identify constitutive promoters (Fig. 3), are less helpful for the selection of TetR- and LacI-repressed promoters because the number of spacer nucleotides that separate operator sites from critical promoter sequences, such as the −10 hexamer, cannot be easily randomized using standard oligonucleotide synthesis. In this case, we took a more directed approach and placed operator sites in altering positions within the core and the proximal region of P2, which were found to be tolerant for sequence variations (Fig. 3b, d). The sequences of five promoters for TetR (PT) and five promoters for LacI (PL) are shown in Fig. 4a. PT and PL promoters were cloned in front of luc and gfp into pPEP8T and pPEP8L, and plasmids were transformed into S. pneumoniae D39. Transformants were grown and analyzed in plate reader assays in duplicate in a high throughput manner, with one well serving as a control for full promoter repression (no inducer), and one well containing saturating concentrations of inducer molecules for maximum induction. Promising candidates were re-streaked, −80°C stocks were prepared, and plate reader experiments were conducted starting from pre-cultured populations (Fig. 4b, c).

Within the PT promoter series, a single tetO site (tetO1) was placed either into the core region (PT1-1) or into the proximal region (PT4-1) of P2, which gave rise to similar results, with expression values for the induced and the repressed state within approximately three orders of magnitude (Fig. 4a, b). For PT5-3, two operator sites were placed both into the core and into the proximal region, resulting in a gene expression activity of $3.4 \times 10^6$ RLU OD$^{-1}$ in the presence of ATc, and $3.6 \times 10^2$ RLU OD$^{-1}$ in the absence of ATc (indistinguishable from control cultures without luciferase), and thus giving rise to a dynamic expression range of approximately four orders of magnitude.
We furthermore analyzed −10 sequence variants of PT4-1, harboring TAAAAT in the case of PT8-2, and TATCAT in the case of PT8-6. Both the induced and the repressed expression values of PT8-2 were downshifted one order of magnitude as compared to PT4-1, demonstrating that promoter leakiness in the repressed state can be reduced by decreasing the overall expression strength (Fig. 4b). PT8-6 showed even lower expression in the induced state as compared to PT8-2; luminescence of the repressed state, however, did not decrease any further because the lower detection limit for luciferase expression was already reached (Fig. 4b). The expression curve of an induction series of PT8-2, as compared to PT4-1, was found to be similar but downshifted in the y-coordinate (Fig. 4d). Interestingly, PT5-3, harboring two tetO sites, showed a more hypersensitive dose-response relationship for ATc induction as compared to PT4-1 and PT8-2 that harbor only one tetO site (Fig. 4d).

For the PL promoter series, the positioning of lacO sites (lacO_sym)^47 within the core region (PL1-2) resulted in a very weak expression of the induced state, presumably because of the introduction of a −35 sequence variation (Fig. 4a, c). In contrast, positioning in the proximal region (PL14-2) resulted in a strong expression of 7.1×10^6 RLU OD⁻¹ when adding IPTG. Luminescence of repressed cultures was not completely suppressed and gave rise to a measurement of 8.1×10^3 RLU OD⁻¹ (Fig. 4c). LacI repressors are known to be able to tetramerize by binding to two operator sites that are in proximity to one another^48. The spacing of these two lacO sites is critical because repressor molecules need to face each other in order to loop DNA and form stable LacI tetramer–DNA complexes^49. Placing an additional lacO site (lacO_1)⁵⁰ at the distal region (72.5 bps upstream of the first operator site, as measured from center to center of the lacO sites) did not lead to an improved repression, and luminescence levels were similar as compared to PL14-2 that harbors only one lacO site. However, when shifting the second operator four bps further upstream, a 8–fold reduction of luminescence in the repressed state was observed as compared to PL14-2; unfortunately, this also led to a 2-fold reduction in the induced state (Fig. 4c). Serendipitously, another clone with a double lacO site was isolated, called PL8-2, which differs from PL8-1 by two insertions, one additional nucleotide in the distal operator site and one additional nucleotide in the proximal region (Fig. 4a). PL8-2 showed the largest induction range of the PL promoter series, spanning four orders of magnitude, with an expression strength of 9.5×10^6 RLU OD⁻¹ in the induced state and 1.2×10^³ RLU OD⁻¹ in the repressed state (Fig. 4c). PL8-2 also showed the strongest hypersensitive response towards IPTG induction (Fig. 4e).

To examine controllable gene expression at the single-cell level, strains harboring the promoters with the highest dynamic range (promoters PT5-3 and PL8-2, strains D-
PEP8T5-3 and D-PEP8L8-2) were grown at varying concentrations of ATc and IPTG, and GFP expression was visualized by fluorescence microscopy. A clear dose–response relationship was found within the induction series of each strain (Fig. 4f, g). Together, these results suggest that PT5-3 and PL8-2, containing orthogonal transcription factors that are regulated by small diffusible molecules, with their large dynamic induction range of up to four orders of magnitude, are the best controllable promoters currently available for *S. pneumoniae*.

**Logic gate and double-inducible system construction**

TetR- and LacI-repressed promoters allow for the external regulation of gene expression via the addition of inducer molecules. Furthermore, these promoters can be used to respond to cell-internal cues. TetR, for example, could be expressed from a promoter that becomes active during competence development, and competent cells would thus shut down the PT5-3 driven expression of genes of interest. Such a regulatory network describes an inverter. To build gene regulatory networks, individual parts need to work in a robust manner, independent of their genomic location. To test whether the here-described repressor systems can be expressed from an ectopic locus, and thus distal to the promoters which they regulate, we integrated the TetR and LacI expression cassettes into the non-essential *prs1* locus (see Methods) (Fig. 5b). Genes of interest under the control of PT5-3 or PL8-2 were integrated at the *amiF* locus via the pPEP9 series of plasmids, which differs from pPEP8T and pPEP8L by omitting the repressor cassettes. Gene expression activity of PT5-3 (strain D-T-PEP9T5-3) and PL8-2 (strain D-L-PEP9L8-2) was similar as compared to strains harboring the repressor cassette at the *amiF* locus (strains D-PEP8T5-3 and D-PEP8L8-2, respectively) (Figs 4d, e and 5c, d).

More complex programming of gene expression, as for example patterns that enable Boolean logic networks, requires the integration of multiple signals into one single response. We wondered whether a P2-based promoter could be used for such a combinatorial regulation, in specific to create a logic AND gate for external induction (ATc and IPTG); note that such a gate describes at the same time a logic NOR gate for cell-internal cues (neither TetR nor LacI). To test this, the synthetic promoter PLT2-2 was constructed that contains a *tetO* site within the core region and a *lacO* site within the proximal region (Fig. 5a). Strain D-LT-PEP9LT2-2, which drives *luc* from PLT2-2 and expresses both LacI and TetR from the *prs1* locus, was found to require both ATc and IPTG to express luciferase (Fig. 5c, d). The absence of ATc, and thus TetR repression alone, was sufficient to decrease luminescence below the detection limit (Fig. 5c). In contrast, LacI repression on its own (with ATc but without IPTG) was not sufficient to
Figure 5 | Construction of an inverter, amplifier and a logic AND gate. (a) Sequence of TetR- and LacI-repressed promoters, with TetR operator sequences in light blue and LacI operator sequences in magenta; transcription start sites are shown in bold, restriction enzyme sites are indicated in dark grey, sequence deviations are underlined. (b) Schematic representation of gene expression regulation constructs, with the regulators expressed from the prs1 locus and the genes of interest expressed from the amiF (pPEP) locus; genR, gentamicin resistance marker; sptR, spectinomycin resistance marker; grey circles indicate transcription terminators. (c, d) Induction series of PT promoters with ATc (c) and of PL promoters with IPTG (d) of the strains shown in b, measured by luminescence (for PLT2-2 induction series}
with ATc (c) 1000 µM IPTG was added to de-repress LacI, and for PLT2-2 induction series with IPTG (d) 100 ng ml⁻¹ ATc was added to de-repress TetR; error bars and fit curves see Methods. (e) IPTG induction series of D-L-PEPdTI (upper scheme, data in the y-coordinate) and D-L-PEPdIL (scheme to the right, data in the x-coordinate), whereby luminescence of D-L-PEPdIL is proportional to the TetR concentration of D-L-PEPdTI; error bars and fit curve see Methods. (f) ATc induction series of D-T-PEPdIL (upper scheme, data in the y-coordinate) and D-T-PEPdT (scheme to the right, data in the x-coordinate), whereby luminescence of D-T-PEPdT is proportional to the LacI concentration of D-T-PEPdIL; error bars and fit curve see Methods.

completely shut down PLT2-2 activity (Fig. 5d). The successful construction of this double-regulated system emphasizes the modularity of prokaryotic promoters; it furthermore provides a framework for the construction of semisynthetic promoters in S. pneumoniae, which respond to their original regulators, and additionally to orthogonal regulators, such as TetR or LacI.

The results above show that both TetR and LacI can be functional within the same cell. Next, we wondered whether the ATc- and IPTG-inducible systems could be used in parallel and independent from one another. We therefore generated the double-inducible integration plasmid pPEPd, with promoter PTdI (identical to PT5-3) positioned in the BglBrick cloning site, and promoter PLdI in an upstream inverse positioning within the terminator insulated BglBrick transfer site¹⁸; PLdI is identical to PL8-2 with the exception of the flanking restriction sites, harboring NheI upstream and BclI downstream of the promoter sequence (Fig. 5a). Two variants of pPEPd were tested (in strains expressing both LacI and TetR at the prs1 locus): D-LT-PEPdT drives luc and gfp from PTdI, and D-LT-PEPdIL drives luc and gfp from PLdI (Fig. 5b). PTdI induction series were found to closely match the ones obtained with PT5-3, without any observable interference from the additionally present LacI (Fig. 5c). PLdI expression, in contrast, deviated from the corresponding PL8-2 pattern, with a 2-fold decrease of the maximum luminescence in the presence of high concentrations of IPTG (Fig. 5d). This decrease, however, could also be observed in a strain that did not express TetR (D-L-PEPdIL; data not shown). Weaker luminescence signals from PLdI, as compared to PL8-2, likely originated from a decreased promoter activity caused by the inverse reading orientation (into the direction of DNA replication) or by the sequence deviation in the proximal region (BclI instead of the BglII site). Alternatively, the translation efficiency might be decreased because of the alteration within the 5’ UTR. Nevertheless, this double-inducible system showed to work without interference between the two systems.

Next, the double-inducible system was used to characterize the TetR and LacI regulated promoters PTdI and PLdI in more detail. There is a fundamental difference
between an external induction and the regulation via cell-internal cues. For external regulation, the amount of inducer molecules determines the expression strength (corresponding to the amount of de-repression), while internal regulation relies on the amount of accumulated repressor proteins. With our double-inducible system, a cell-internal regulation could be simulated by controlling the amount of TetR via PLdi induction (strain D-L-PEPdiTT; upper scheme in Fig. 5e), or by controlling the amount of LacI via PTdi induction (strain D-T-PEPdiLL; upper scheme in Fig. 5f). Repressor expression of these strains, at a specific induction level, can furthermore be matched to luciferase expression of corresponding strains (D-L-PEPdiL and D-T-PEPdiT, respectively; schemes to the right in Fig. 5e, f), and thus be displayed as RLU OD$^{-1}$ equivalents (Fig. 5e, f). Interestingly, TetR repression of PTdi was found to be more hypersensitive than LacI repression of PLdi, with a repressor concentration range from maximum promoter activity to full repression within two orders of magnitude for TetR, as compared to three orders of magnitude for LacI. PTdi showed the same maximum expression in the absence of TetR as in the presence of fully de-repressed TetR with ATc. In contrast, in the case of PLdi, luminescence was 2–fold higher in the absence of LacI as compared to fully de-repressed LacI with IPTG. This observation confirms the findings of a previous study with *E. coli* that showed that de-repressed LacI does not completely lose its affinity for *lac*O sites$^{54}$. Note that the regulatory networks assembled in pPEPdiTT and pPEPdiLL describe amplifiers; they could also be used to increase cell-internal signals.

**Construction and characterization of toggle switches**

*S. pneumoniae* is one of the leading bacterial threats worldwide, and there is evidence that the heterogeneous expression of virulence factors within bacterial populations (such as the polysaccharide capsule) contributes importantly to its pathogenicity$^{9,20,55}$. To be able to improve our understanding of such expression patterns, we aimed at engineering synthetic regulatory networks that are bistable, and that can thus give rise to bimodal population distributions$^{11}$. A classic example of such a network is the so-called toggle switch, in which two transcription regulators repress the expression of each other$^{13}$. The *prs1* site was used to integrate a genetic toggle switch into the pneumococcal genome (Fig. 6a). Note that this represents, to our knowledge, the first description of a single-copy chromosomally integrated synthetic toggle switch, in contrast to existing toggle switches that were constructed on replicating plasmids. Toggle switch 1 (TS1) harbors the promoters PTts1 (identical to PT5-3; restriction sites *Xba*I upstream and *Ase*I downstream) driving *lacI*, and PLts1 (identical to PL8-2; restriction sites *Xba*I upstream and *Bgl*II downstream) driving *tetR* (Fig. 6a). When assuming that PTts1 and PLts1 result in identical expression patterns
as PTdi and PLdi, then the TetR regulated expression of PTdi (Fig. 5e) and the LacI regulated expression of PLdi (Fig. 5f) can be combined into one plot, displaying the gene expression regulation of the toggle switch (Fig. 6c). Intersections of the two expression curves (black circles) represent steady states in which gene expression activity of the two promoters and the concentration of the two repressors are in equilibrium. Three steady states emerged, with the outer two representing stable states, and the middle one (asterisk) representing the threshold of the bistable switch. The proximity of this threshold to one of the stable states indicates that naturally occurring fluctuations within cells might be sufficient to trigger switching events. Fluctuations can originate, amongst others, from unequal distributions of repressor molecules after cell division, or from prolonged promoter accessibility after spontaneous repressor dissociation54,56,57.

In the case of the promoter pair PTts1 and PLts1 (TS1), the predicted positioning of the steady states indicates that cells containing high levels of TetR and low levels of LacI are more likely to remain in their current state than cells containing low levels of TetR and high levels of LacI (Fig. 6c). We therefore created two additional toggle switches (TS2 and TS3) by modifying the −10 sequence of PLts1 (TS2, PLts2 −10 sequence TAAAAT; TS3, PLts3 −10 sequence TATCAT) with the goal of finding bistability patterns with similar spacing between the threshold and the two stable states. Deviations of the canonical −10 sequence were previously shown to downshift the expression curve of an induction series (10–fold for TAAAAT, and presumably 20–fold for TATCAT) (Fig. 4b, d). A hypothetical expression curve was drawn into Fig. 6c (dashed line) to help indicating the shift of steady states (grey circles) of these newly created toggle switch variants.

To characterize the putative toggle switches, reporter genes were integrated into the S. pneumoniae D39 genome at the amiF locus (Fig. 6a). On the basis of pPEPdi, pPEP10 was created, with the erythromycin resistance marker erm, luc and gfp driven from PTdi, and the kanamycin resistance marker aphA driven from PLdi (Fig. 6a). Toggle switch strains were triggered, either with IPTG or with ATc, by plating and overnight incubation, followed by 8 h cultivation in liquid medium in the presence of inducer (to allow for the establishment of stable expression equilibria; Fig. 6b). Next, induced cultures were re-plated and re-grown for 8 h in liquid medium without inducer to allow for the settlement of gene expression at stable states, and for switching events to occur (Fig. 6b).

Cells that had experienced the describe treatment were analyzed for resistance towards: (i) spectinomycin (resistance marker of PEP), indicating the total number of viable cells; (ii) kanamycin, indicating the number of cells with low levels of LacI (and presumably high levels of TetR; from here on i-state, for IPTG pretreatment); (iii) erythromycin, indicating the number of cells with low levels of TetR (and presumably high levels of LacI; from here
Figure 6 | Transcriptional toggle switches. (a) Schematic representation of strain D-TS-PEP10 containing a transcriptional toggle switch at the prs1 locus and genes of interest at the amiF locus; genR, gentamicin resistance marker; spfR, spectinomycin resistance marker; aphA, kanamycin resistance marker; erm, erythromycin resistance marker; grey circles indicate transcription terminators. (b) Work flow of D-TS-PEP10 induction and the subsequent identification of switching events; IPTG, 1000 µM; ATc, 100 ng mL⁻¹;
SYNTHETIC BIOLOGY TOOLS TO CONTROL GENE EXPRESSION IN STREPTOCOCCUS PNEUMONIAE

on a-state, for ATc pretreatment); (iv) kanamycin and erythromycin, serving as a control (Fig. 6d–f). Cells of colonies that emerged in the presence of both kanamycin and erythromycin likely contained mutations. Cell populations were furthermore analyzed by flow cytometry, whereat high GFP-fluorescence indicated cells in the a-state, and low fluorescence indicated cells in the i-state (Fig. 6d–f). Additionally, cells of both induced cultures and of cultures after re-plating without inducer were analyzed by fluorescence microscopy (Fig. 6d–f).

Cells harboring TS1 that were previously treated with IPTG were all found in the corresponding i-state (Fig. 6d). In contrast, only approximately half of the TS1 cells from ATc pretreatment were found in the a-state, and the remaining half had switched to the i-state (Fig. 6d). These findings matched the prediction made on the basis of the plot in Figure 6c. In the case of TS2, cells from IPTG-pretreated cultures were also all found in the i-state (Fig. 6e). For ATc pretreatment, the number of TS2 cells that had switched from the a-state to the i-state was found to be two orders of magnitude lower compared with TS1, with approximately one switching event per 200 cells (Fig. 6e). Transitions from the a-state to the i-state were found to occur even less frequently in TS3 cultures, with only one out of 10000 cells showing kanamycin resistance (Fig. 6f). However, in TS3 strains, also cells of cultures that were pretreated with IPTG were found to be able to switch, from the i-state to the a-state, with a frequency of approximately one out of 1000 cells within the observed time (Fig. 6f).

Remarkably, an additional prediction that was made on the basis of the plot in Figure 6c could be confirmed. By decreasing the promoter strength of PLts, the a-state of the toggle...
switch would be expected to correspond to a lower TetR concentration, which in turn should result in cells expressing higher levels of GFP. Indeed, flow cytometry measurements showed that the peak of the fluorescence intensity shifted in direction of the $x$-coordinate, from TS1 over TS2 to TS3, with the mode of TS1 cells found at output level #362, for TS2 cells at #388, and for TS3 cells at #398 (arbitrary units, see also Methods) (Fig. 6d–f).

**Discussion**

Here, we created and characterized a selection tool for gene expression in *S. pneumoniae* by combining the counterselection marker *pheS*, the selection marker *erm*, the population reporter *luc* and the single-cell reporter *gfp*. This selection platform was used to analyze libraries of constitutive pneumococcal promoters. Operator sequences for the *E. coli* derived repressors TetR and LacI were integrated into core and proximal promoter regions that showed tolerance for sequence variations, and a set of inducible promoters was created and characterized. On the basis of the two promoters with the largest dynamic range, PT5-3 and PL8-2 spanning approximately four orders of magnitude, regulatory networks of higher complexity were successfully assembled and analyzed. These networks include inverters, an AND gate, amplifiers and toggle switch variants; to our knowledge, this is the first report of a single-copy genomic integrated synthetic toggle switch.

*S. pneumoniae* represents an interesting candidate for synthetic biology applications. The straightforward uptake of exogenous DNA and stable integration into the genome via natural competence allows for stable copy numbers and tightly controlled gene expression of artificial regulatory networks. Furthermore, pneumococci have a small genome of approximately 2 million bps (and thus reduced genetic redundancy), and over 10000 genome sequences are available, ranging from strains used for fundamental research, isolates from healthy carriers and clinical isolates from patients with invasive diseases. Providing tools for fundamental research questions might represent a starting point for real applications of synthetic biology. It is therefore of importance to adapt tools developed for model organisms, as a proof of principle, to bacteria of interest, such as human pathogens, where these methods might help in solving research questions that are difficult to answer with conventional methods.

To this effect, the synthetic biology approaches pioneered here allowed us to integrated complex gene regulatory networks, such as genetic toggle switches, into the genome of pneumococcus to accurately control gene expression. These networks could, in a next step, be used to drive virulence factors, and thus allow for *in vivo* investigations of the contribution of gene expression regulation on the pathogenicity of the important human
pathogen *S. pneumoniae*. In addition, the approaches used here may serve as an example for synthetic biology projects in unrelated organisms.

**Methods**

**Strains and growth conditions**

*S. pneumoniae* D39 was used throughout, and *E. coli* MC1061 was used for sub-cloning. *E. coli* competent cells were obtained by CaCl$_2$ treatment$^{59}$; transformations were carried out via heat-shock at 42°C. *S. pneumoniae* transformations were carried out with cultures at OD (600 nm) 0.1 in the presence of 1 ng ml$^{-1}$ CSP (competence-stimulating peptide)$^{60}$. Promoters and genes of interest were assembled in pPEP$^{18}$ via BglBrick cloning$^{19}$ followed by integration into the D39 genome at the amiF locus (by replacing a non-essential fragment spanning base pairs 1678506 to 1678907 according to the D39 genome; GenBank NC_008533)$^{61}$. Integration constructs inside the prsI locus$^{51}$ (replacing base pairs 29751 to 30077) were assembled via Gibson assembly$^{62}$ and directly transformed to *S. pneumoniae*.$^{63}$ Pneumococcal cells were cultivated in C+Y medium$^{64}$ (pH 6.8) supplemented with 0.5 µg/ml D-luciferin for luminescence measurements, at a temperature of 37°C. Pre-cultures for all experiments were obtained by a standardized protocol, in which previously exponentially growing cells from −80°C stocks were diluted to OD 0.005 and grown until OD 0.1 in a volume of 2 ml medium in tubes that allow for direct (in tube) OD measurement. For selection assays, and to determine the number of colony forming units, cells were plated inside Columbia agar supplemented with 3% (v v$^{-1}$) sheep blood and incubated overnight at 37°C.

**Microtiter plate reader assays**

Costar 96 well plates (white, clear bottom) with a total assay volume of 300 µl per well were inoculated to the designated starting OD value. Microtiter plate reader experiments were performed using a TECAN infinite pro 200 (Tecan Group) by measuring every 10 min with the following protocol: 5 s shaking, OD (595 nm) measurement with 25 flashes, luminescence (RLU, relative luminescence units, a.u.) measurement with an integration time of 1 s. Average and s.e.m. of normalized luminescence (RLU OD$^{-1}$) were determined between OD 0.01 and OD 0.02, on the basis of three measurements of duplicates (and thus six data points). Fit curves for induction series were based on the 4 parameter Hill equation:

$$f(x) = s + \frac{\beta}{\left( \frac{K_A}{x} \right)^n + 1}$$
with $s$, minimum expression; $\beta$, maximum expression; $n$, Hill coefficient; $K_a$, dissociation constant. The following parameters were found ($s; \beta; n; K_a$): Fig. 4d: D-PEP8PT4-1 (4000; $6.3 \times 10^6$; 2.1; 6.0), D-PEP8PT8-2 (360; $7.1 \times 10^5$; 2.1; 9.3), D-PEP8PT5-3 (360; $3.4 \times 10^6$; 3.3; 11); Fig. 4e: D-PEP8PL14-2 (8100; $7.1 \times 10^6$; 1.8; 54), D-PEP8PL8-1 (1000; $2.9 \times 10^6$; 1.9; 140), D-PEP8PL8-2 (1200; $9.5 \times 10^6$; 2.2; 54); Fig. 5c: D-T-PEP9PT5-3 (340; $3.4 \times 10^6$; 3.3; 15), D-LT-PEP9PLT2-2 (350; $8.4 \times 10^5$; 2.1; 9.7), D-LT-PEP9PTdi (370; $3.4 \times 10^6$; 3.3; 13); Fig. 5d: D-L-PEP9PL8-2 (1100; $9.0 \times 10^6$; 2.2; 63), D-LT-PEP9PLT2-2 (1800; $9.0 \times 10^5$; 1.6; 62), D-LT-PEP9PLdi (870; $4.4 \times 10^6$; 2.0; 65); Fig. 5e: D-L-PEPdiTT (380; $3.3 \times 10^6$; −2.9; 7000); Fig. 5f: D-T-PEPdiLL (870; $1.0 \times 10^7$; −1.9; 22000). Note that ATc was annotated in the commonly used form of ng ml$^{-1}$, which corresponds to 2.16 nM (the molecular weight of anhydrotetracycline hydrochloride is 462.88 g mol$^{-1}$).

**Microscopy**

A Nikon Ti-E microscope equipped with a CoolSnapHQ2 camera and an Intensilight light source were used. Microscopy was carried out by spotting cells on a 10% polyacrylamide slide containing PBS, inside a Gene Frame (Thermo Fisher Scientific) that was sealed with the cover glass to guaranty stable conditions. Images of fluorescing cells were taken with the following protocol and filter settings: 0.3 s exposure for phase contrast, 1 s exposure for fluorescence at 440−490 nm excitation via a dichroic mirror of 495 nm and an emission filter at 500−505 nm. For the induction series shown in Figure 4e, f, D-PEP8T5-3 and D-PEP8L8-2 cells were pre-cultivated according to standard protocol (to OD 0.1), followed by a 100–fold dilution and regrowth to OD 0.1 in the presence of ATc and IPTG, respectively.

**Flow cytometry**

The fluorescence intensity of $10^4$ cells was measured by a BD FACS Canto Flow Cytometer (BD Bioscience) at medium flow, with the detectors for forward scatter and side scatter set to 200 V and 500 V, respectively. A gate was defined on the basis of forward and side scatter measurements to exclude the recording of particles that deviated from normal S. pneumoniae cells. The detector for fluorescence was set to 750 V. Results shown in Figure 6d–f represent smoothed data (running average) of output levels #0 to #600 of a 10–bit channel (total number of 1024 output levels), whereas measurements from output levels #0 to #300 were fitted to a normal distribution for visual clarity; the detection of weakly fluorescing cells in our flow cytometer suffered from machine biases, showing periodically repeating stretches of empty reads and culmination at output level #0.
Acknowledgements
We thank Melinde Wijers (University of Groningen, undergraduate program) for excellent technical assistance during the initial phase of this project.

Author Contributions
R.A.S and J.-W.V. initiated the project. R.A.S. performed all experiments. R.A.S. and J.-W.V. wrote the manuscript.

References


