



CRISPRi-seq for genome-wide fitness quantification in bacteria

Vincent de Bakker^{1,3}, Xue Liu^{1,2,3}, Afonso M. Bravo¹ and Jan-Willem Veening¹✉

CRISPR interference (CRISPRi) is a powerful tool to link essential and nonessential genes to specific phenotypes and to explore their functions. Here we describe a protocol for CRISPRi screenings to assess genome-wide gene fitness in a single sequencing step (CRISPRi-seq). We demonstrate the use of the protocol in *Streptococcus pneumoniae*, an important human pathogen; however, the protocol can easily be adapted for use in other organisms. The protocol includes a pipeline for single-guide RNA library design, workflows for pooled CRISPRi library construction, growth assays and sequencing steps, a read analysis tool (2FAST2Q) and instructions for fitness quantification. We describe how to make an IPTG-inducible system with small libraries that are easy to handle and cost-effective and overcome bottleneck issues, which can be a problem when using similar, transposon mutagenesis-based methods. Ultimately, the procedure yields a fitness score per single-guide RNA target for any given growth condition. A genome-wide screening can be finished in 1 week with a constructed library. Data analysis and follow-up confirmation experiments can be completed in another 2-3 weeks.

Introduction

Linking genotype to phenotype is a major goal in biology. To this end, uncovering the functions of genes is vital. Genome-wide gene function studies in bacteria have thus far mainly employed massively parallel sequencing combined with transposon mutagenesis (e.g., transposon sequencing (Tn-seq), transposon-directed insertion site sequencing (traDIS), high-throughput insertion tracking by deep sequencing (HITS), insertion sequencing (INSeq))¹. However, these approaches are not suitable for studying essential genes, as insertion mutants of essential genes cannot be constructed. Alternatively, CRISPR interference (CRISPRi) has been used widely to study gene functions in bacteria^{2–9}.

CRISPRi is a method to repress expression of a target gene⁹. A single-guide RNA (sgRNA) leads a catalytically inactive (dead) Cas9 (dCas9) protein to a site on the genome with a sequence complementary to its spacer sequence next to a protospacer adjacent motif (PAM)^{2,3}. When bound to the target gene, dCas9 serves as a roadblock for RNA polymerase, thereby blocking transcription^{2,3}. Targeting a gene with an sgRNA on the nontemplate strand close to the transcription start site (TSS) typically efficiently represses transcription initiation or transcription elongation^{2,3,6}. sgRNAs can be designed to target any open reading frame of interest that harbors an appropriate PAM, and the timing and severity of the knockdown can be controlled with inducible dCas9 promoters^{5,7}. As such, all genes of a given genome can be studied with appropriately designed sgRNA libraries.

In general, two strategies can be adopted for genome-wide screenings with CRISPRi. One is with an arrayed library, in which each mutant carrying one sgRNA is handled individually. The other is with a pooled library, in which multiple mutants are grown together in a pooled fashion. The arrayed library allows for direct observation of diverse phenotypes such as growth defects, cell morphology changes or luciferase/fluorescence activity^{5,7,10}. However, it is laborious and time-consuming to perform genome-wide screenings with arrayed libraries, since each strain has to be processed individually. In contrast, screening with pooled CRISPRi libraries is more convenient but allows for limited phenotype measurements besides gene fitness. Indeed, deep sequencing can reveal changes in sgRNA abundancies of the pooled library upon selective pressures, reflecting the conditional fitness of their targets. In fact, pooled CRISPRi libraries combined with next-generation sequencing

¹Department of Fundamental Microbiology, Faculty of Biology and Medicine, University of Lausanne, Lausanne, Switzerland.

²Guangdong Key Laboratory for Genome Stability and Human Disease Prevention, Department of Pharmacology, International Cancer Center, Shenzhen University Health Science Center, Shenzhen, China. ³These authors contributed equally: Vincent de Bakker, Xue Liu.

✉e-mail: jan-willem.veening@unil.ch

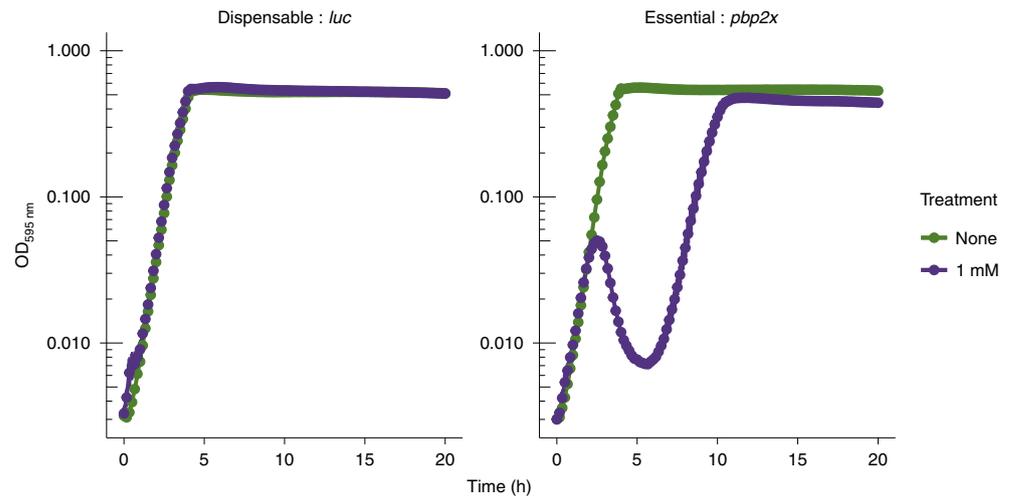


Fig. 1 | Growth curves of individual CRISPRi strains indicate essentiality of sgRNA targets. *S. pneumoniae* D39V strains harboring an sgRNA targeting firefly luciferase (*luc*) or penicillin-binding protein 2x (*pbp2x*) were grown with (1 mM) or without IPTG as inducer of dCas9 expression. Cell density was measured every 10 min as OD_{595 nm} in five technical triplicates, of which averages and standard errors are depicted. Bacteria containing an sgRNA targeting an essential gene (here *pbp2x*) cannot grow upon CRISPRi induction as the expression of this gene is repressed, whereas no growth defect is observed for those bacteria with an sgRNA targeting a dispensable gene (here *luc*). Strains in which an essential gene is repressed often display the emergence of loss-of-function mutations in the CRISPRi system, allowing growth to resume after a couple of hours, as can be observed here for the *pbp2x* knockdown strain. Strains XL25 and XL28 were obtained from and grown as presented in ref. ⁵.

(CRISPRi-seq) can be used to study gene fitness in bacteria under different conditions on a genome-wide level, both in vitro and in vivo^{4,6,11–18}.

Development of the protocol

Previously, we made an IPTG-inducible, titratable CRISPRi system in *Streptococcus pneumoniae* D39V (genome accession CP027540)¹⁹ to study essential gene function in vitro in an arrayed screening fashion (Fig. 1)⁵. Recently, we developed a doxycycline (dox)-inducible system, which is also applicable in vivo¹¹. We combined the latter system with a small but genome-wide sgRNA library (1,499 sgRNAs targeting in total 2,111 genetic features), and pooled growth assays and Illumina sequencing to establish the CRISPRi-seq method for whole-genome gene fitness quantification. In this protocol, we describe in detail how to perform CRISPRi-seq, from sgRNA design and CRISPRi library construction to downstream analyses such as gene fitness quantification (Fig. 2). For clarity, we focus on the IPTG-inducible system in this protocol, but the dox-inducible system can be constructed in a similar way.

Applications of the protocol

This protocol was designed for usage in *S. pneumoniae* D39V. However, the system is easily transplantable into other pneumococcal strains. In addition, the ready-made sgRNA pool (Addgene #170432) we describe here also covers core operons of other commonly studied pneumococcal strains. We previously estimated a coverage of 90.3% of all annotated genetic elements for strain R6, 72.2% for TIGR4, 66.6% for Hungary 19A-6, 71.6% for Taiwan 19F-14, 66.8% for 11A and 72.1% for G54, despite the library being designed for *S. pneumoniae* D39V (ref. ¹¹). Therefore, application of CRISPRi-seq to such strains should be straightforward using the protocol presented here.

In addition, similar systems can be developed for other bacterial species, provided they are transformable, using guidelines provided here. Pooled CRISPRi screenings have been performed for a wide range of model organisms, including *Escherichia coli*^{16,15,16,20,21}, *Bacillus subtilis*²¹, *Staphylococcus aureus*²², *Mycobacterium tuberculosis*^{12,23} and *Vibrio natriegens*⁴. This demonstrates the possibility of establishing a CRISPRi-seq system similar to ours in these species as well, especially with small, practical sgRNA libraries as described here.

We mainly focus on the fitness quantification of target genes in this protocol. However, the methodology can easily be extended to explore more complex biological questions. CRISPRi-seq

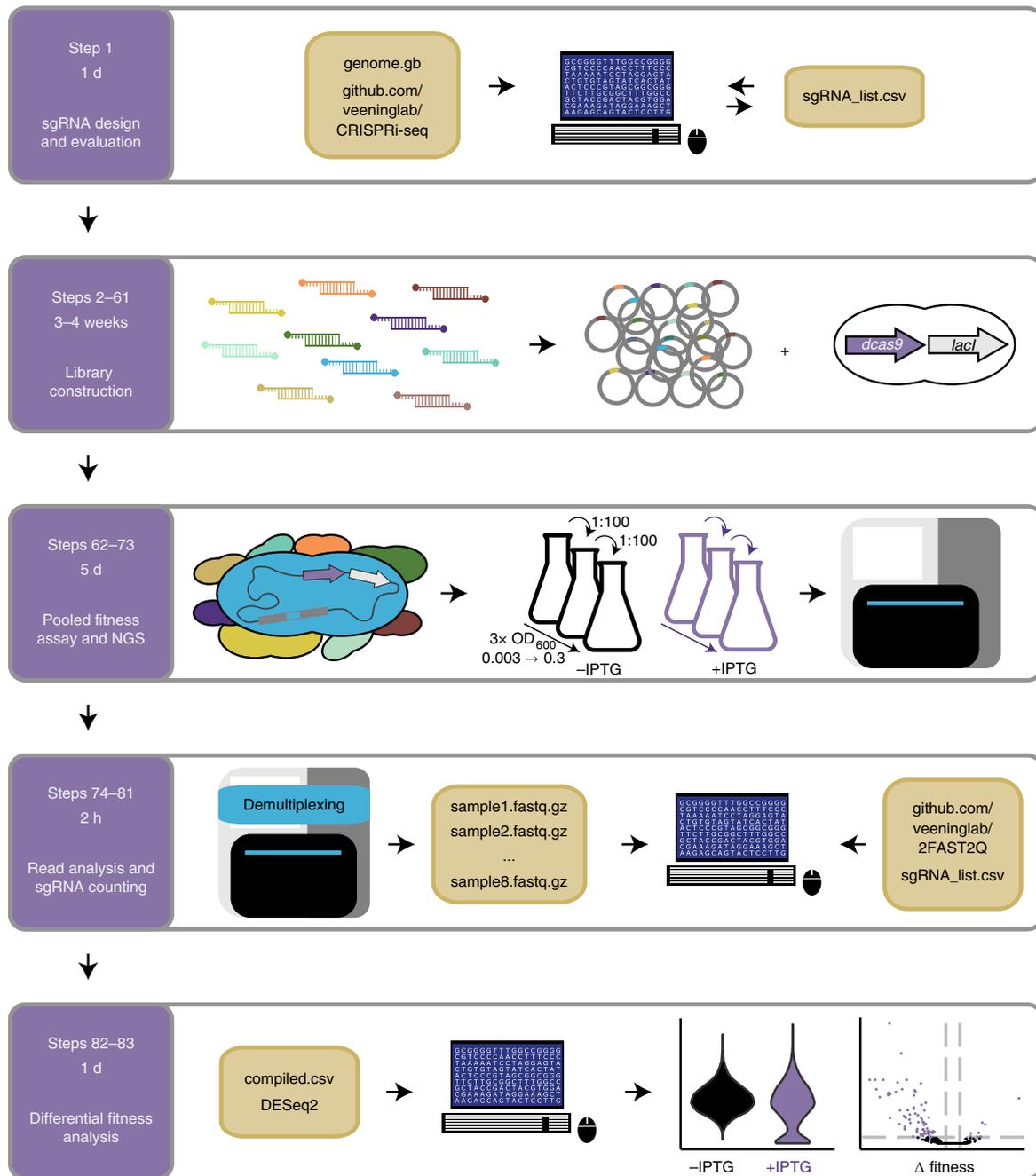


Fig. 2 | Overview of the CRISPRi-seq workflow. The major phases of the protocol are delineated, including the steps and times involved as described in the Procedure. A suitable sgRNA library is selected or designed prior to the construction of the corresponding sgRNA plasmid pool. This pool is integrated into the chromosome of a constructed host strain harboring an inducible dCas9 CRISPRi system, yielding the CRISPRi library. This library is grown in a pooled fashion under a selection pressure of interest either with (purple) or without (black) the inducer. Induction during more generations while maintaining exponential growth is achieved through back-dilutions. Subsequently, the sgRNAs are extracted from the gDNA pool and subjected to Illumina sequencing, resulting in one .fastq.gz file per sample. These files are processed with the new tool 2FAST2Q, which returns the quality-checked sgRNA counts per sample. sgRNA depletion or enrichment in induced versus non-induced samples represents target fitness, which is here analyzed with the R package DESeq2 (ref. ⁵¹). Figure adapted from ref. ¹¹.

screenings performed under conditions with antibiotic stresses can identify genes related to sensitivity towards antibiotics²². Gene fitness comparisons between wild-type and specific mutant backgrounds, or dual-sgRNA studies, can uncover gene–gene interactions, such as synthetic lethality or functional redundancy of gene pairs, which facilitates gene function studies^{22,24–27}. Uninduced pooled libraries can be used as neutral barcodes to establish bottlenecks under selective conditions¹¹. Lastly, CRISPRi-seq can reveal important genes during host–microbe interactions in animal infection models¹¹.

Comparison with other methods

Tn-seq approaches serve a similar goal, but are by definition confined to a genome-wide scale and biased towards larger genes because of the random insertion property of transposons²⁸. Although Tn-seq libraries are generally faster and cheaper to build, they tend to require thousands of mutants to reach saturation, increasing the needed sequencing capacity for full library characterization. In addition, polar effects of Tn insertions can complicate data interpretation²⁹. It is also noteworthy to mention that the Tn-seq amplicon preparation for Illumina sequencing is typically more laborious and error-prone than the here-presented single-step PCR protocol for sequencing pooled sgRNA libraries^{11,30}. Furthermore, in conditions with stochastic bottlenecks, such as in murine models of infection in which a random part of the library dies off due to factors such as host immune system clearance or limited nutrient availability, the use of large transposon libraries is problematic for determining whether a given mutant disappears by chance or for being less fit^{11,31}. CRISPRi-seq, on the other hand, is targetable, inducible and titratable. The former allows for the usage of small libraries to target any collection of genes of interest and for controlled polar effects. Small libraries are easier to handle and facilitate downstream analysis. Moreover, CRISPRi can be used to study essential genes, which is not easily possible using Tn-seq³², as by definition, essential genes have few or no transposon insertions. In addition, expression repression activity can be fine-tuned by titrating the inducer or by deliberate design of sgRNA–target mismatches^{21,33}. While Tn-seq methods can reach sub-gene-level resolution in quantifying gene fitness, CRISPRi-seq is bound to the operon level because of downstream effects^{2,3,5,7}. Comparisons between results of these two methods should therefore be made on the operon level. CRISPRi-seq recapitulates the previously Tn-seq-identified essential loci, at both the gene and operon level (Supplementary Fig. 1). Indeed, both Tn-seq datasets and our CRISPRi results agree on the essentiality of 133 operons, reflecting 269 different essential genes. The observed study-specific gene essentiality predictions could arise from differences in growth media, or from the gene essentiality calling pipeline itself. Overall, these data demonstrate the potential of CRISPRi-seq as a valid alternative to Tn-seq, allowing the examination of (non)essential loci fitness with smaller library sizes.

For our setup, we selected one sgRNA with a high specificity and efficiency per operon. This strategy results in a relatively small library size: $\sim 10^3$ for pneumococcal strains. In comparison, previous pooled CRISPRi screenings were carried out using larger libraries ($\sim 10^4$ – 10^5), with multiple sgRNAs targeting each gene^{6,12,34}. On the one hand, these libraries are more expensive for sequencing that reaches the same sequencing depth, due to their size. Additionally, small sgRNA pools are practically easier to handle during experiments in the laboratory, especially for bacterial strains with low transformation efficiency that are challenging to transform using a large sgRNA pool with good coverage. Furthermore, in a small pool, the sgRNAs can be designed more specifically to avoid off-target effects because they are few, and the resulting libraries are more suited for growth conditions with a tight bottleneck, such as animal models¹¹. On the other hand, it is sometimes difficult to estimate (off-target) knockdown efficacy and sgRNAs occasionally show unexpected on- and off-target behavior. This is much less of an issue in a setup in which each gene is targeted by multiple sgRNAs, and the gene fitness effect is subsequently determined by differential depletion of the full set of sgRNAs targeting the gene. Moreover, if operon structures or TSS annotations are either unknown or cryptic, some genes might not be targeted properly by a library designed to target features at the operon level with a single sgRNA. Lastly, since reverse polarity of CRISPRi-mediated knockdown seems to be species-specific¹⁷, it might be worthwhile to design sgRNAs for each feature, rather than for each operon, as this might give sub-operon fitness resolution. Altogether, the use of small libraries can have considerable advantages, partially depending on the experimental conditions of interest. However, the trade-off between efficiency and effectiveness to choose a compact small pool over a large pool should always be carefully evaluated. To meet different requirements of users, here we provide an sgRNA design pipeline that facilitates both strategies.

In contrast to pooled CRISPRi screenings, arrayed CRISPRi screenings have the key advantage that other features besides fitness can be measured, such as cell morphology and fluorescent signals^{5,10,35}. Recently, a combination of imaging within microfluidics devices with a small pooled CRISPRi library was shown to be insightful in coupling effects of transcriptional knockdown of DNA replication genes on the *E. coli* cell cycle³⁶. A main advantage of arrayed or single-cell analysis of individual sgRNA strains is that growth and phenotype upon transcriptional repression are not influenced by the presence and growth of competing strains. Regardless of these disadvantages compared with arrayed libraries, for genome-wide fitness assays as described here, pooled libraries are highly efficient and can

quickly generate testable hypotheses. Additionally, pooled and arrayed libraries can be compatible with each other. An arrayed library can be pooled together after construction, while a pooled library can be sorted into an arrayed library in several fast, cheap and efficient ways^{37,38}.

Expertise needed to implement the protocol

Deep sequencing is required for this protocol. The methods are described for an Illumina MiniSeq sequencing system, but if no comparable system is available to the user, use of a sequencing service or core facility would be needed.

Limitations

Genes without a PAM sequence cannot be targeted by CRISPRi and, as such, cannot be silenced with this method if they do not share an operon with other targetable genes. Moreover, some (small) genes might not be targetable uniquely, for instance if they only have a PAM in certain repeated regions. Sometimes these issues can be circumvented by targeting the corresponding promoters. Although polar effects might indeed allow one to repress otherwise nontargetable genes, this also means that this method measures fitness on the operon level and, thus, in principle does not have gene-level resolution.

Fitness, computed as the \log_2 fold change (\log_2 FC) of sgRNA abundance between experimental and control conditions, can be influenced by the timepoint of sampling. Longer duration of treatment with the inducer leads to larger absolute fold changes, and thus to larger fitness gains or losses. This is beneficial to explore genes with marginal fitness changes. However, the sgRNAs targeting genes that are critical for bacterial growth might drop out from the population after long treatment, resulting in 0 counts after sequencing. As a consequence, the differences among these sgRNAs are concealed. So, to obtain richer information, different timepoints of sampling are necessary in the screen. The fitness determined by CRISPRi-seq should be interpreted carefully, in the context of the timing of sampling.

The pooled growth approach described here is practical, but cannot distinguish between competitive and environmental effects on sgRNA selection. Strains with different sgRNAs can interact when grown together, which might influence sgRNA selection in the sample as a whole (similar to pooled transposon mutant libraries). In addition, sgRNA selection can be determined by bottlenecks, e.g., when applied to in vivo models, even for small libraries like ours¹¹. In any case, we advise using this protocol as a screening method and following up with hit validation experiments, such as growth assays with clean knockout mutants or individual CRISPRi strains.

Experimental design

Setup of the host strain (Steps 41–45)

A functional CRISPRi system requires two key elements: the dCas9 protein and an sgRNA. Here, expression of dCas9 is controlled by an IPTG-inducible system, whereas sgRNAs are constitutively expressed. *S. pneumoniae* D39V with an inducible *dcas9* was constructed as host strain for later CRISPRi library assembly. For the IPTG-inducible library, the parental strain was first transformed with constitutively expressed *lacI* integrated at the *prs1* locus, and then with P_{lac} -*dcas9* at the *bgaA* locus. The CRISPRi library is constructed by subsequently transforming the host strain with a pool of sgRNAs cloned in an integrative vector (Fig. 3d).

sgRNA design and evaluation (Step 1)

The ready-made sgRNA pool (Addgene #170432) used in this protocol was manually designed according to the annotation in *S. pneumoniae* D39V (ref. ¹⁹). Because of the well-documented polar effects of CRISPRi, the sgRNAs were designed on an operon level as much as possible¹⁷. For each of the annotated primary operon TSS of *S. pneumoniae* D39V¹⁹, one sgRNA was selected to target the nontemplate strand with high specificity and in close proximity to the primary operon TSS, since these criteria have been associated with higher repression activity by blocking transcript elongation^{2,3,6}. This covered ~65% of the annotated genes. For the remaining genetic features that lacked primary operon TSS annotation, one sgRNA was selected for each, aiming for high specificity and close proximity to the start codon in an attempt to obstruct elongation early on. This yielded a library of 1,499 sgRNAs covering 2,111 out of 2,146 (98.4%) annotated genetic features for this strain¹¹. Users can either use this existing library for their own experiments and strains, or design a new library. This ready-made sgRNA pool was deposited in Addgene, and the distribution of the sgRNAs

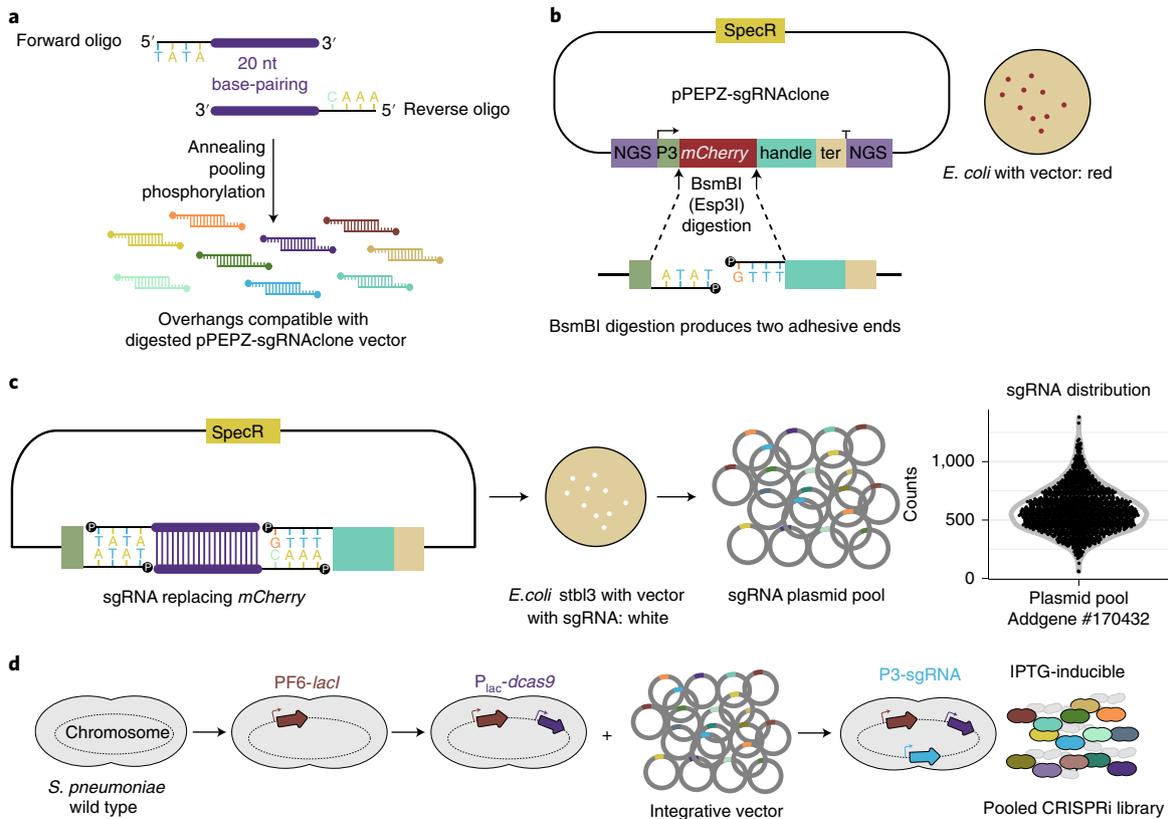


Fig. 3 | Construction of the pooled CRISPRi library. **a**, sgRNA pool construction by annealing oligo pairs with an overhang that is compatible with the backbone vector. **b**, Structure of backbone vector pPEPZ-sgRNAclone with spectinomycin resistance marker (SpecR). The constitutive P3 synthetic promoter-driven *mCherry* gene makes successfully transformed *E. coli* colonies red. The *mCherry* cassette is replaced by the sgRNA after digestion due its overhang-compatible Esp3I/BsmBI restriction sites. The sgRNA is ligated into the vector directly adjacent to sequences encoding its dCas9 handle and terminator (ter) and directly after the P3 promoter. Illumina amplicon elements (NGS) are integrated in the vector, in front of the P3 promoter and after the terminator (ter) of the sgRNA, to facilitate sequencing library construction using a one-step PCR later on in the Procedure. **c**, After sgRNA insertion into the backbone vector, the plasmids are transformed into *E. coli*. Successful transformation is marked by white colonies, due to the replacement of *mCherry*. The sgRNA plasmid pool is subsequently harvested from successful transformants. NGS of the pool should reveal the sgRNA counts are approximately normally distributed, as is shown for our ready-made sgRNA library (points are sgRNAs, gray outline depicts corresponding violin plot). **d**, Construction of a host strain with an IPTG-inducible *dcas9*. Upon transformation with the sgRNA plasmid pool, the CRISPRi strain library is obtained. Both *dcas9* and the sgRNA are integrated into the chromosome in the pneumococcal library. Figure adapted from ref. ¹¹.

in the pool is shown in Fig. 3c. If used, it is important to evaluate the expected efficiency of the ready-made library on the genome of interest.

To that end, all potential binding sites of each of the sgRNAs in the given library on the genome must be determined. This not only allows for a priori genome coverage assessment of the library, but also for a posteriori identification of target features responsible for any observed phenotype (including binding sites with mismatches and potential off-targets). We therefore constructed a computational pipeline to identify all possible binding sites of any sgRNA library on any given bacterial genome (either a local GenBank file or automatically extracted from a National Center for Biotechnology Information database), up to a user-defined cumulative number of mismatches per sub-spacer region. Several studies have reported an inverse correlation between knockdown efficiency and the number of mismatches and their within-spacer proximity to the PAM^{2,3,21,33}. For instance, an sgRNA with two mismatches within a 7 nt PAM-proximal seed region with any locus on the genome is unlikely to effectuate substantial transcriptional repression at that locus, whereas sgRNAs with up to 11 mismatches in the PAM-distal region of the spacer could potentially completely shut down transcription^{2,21,34}. As such, we detect by default for each sgRNA binding sites with maximally 1 mismatch in the first seven nucleotides (PAM-proximal to -distal), 2 mismatches in the first nine nucleotides and 11 mismatches in the complete 20 nt spacer, although these settings can be adjusted by the user. This (mis)matching-based detection of biologically relevant binding sites is carried out by the Python-built Tool for Identification of Nucleotide-wise Dual-sequence, Region-delimited identity

(TINDRi.py) presented here. The tool is called automatically from within the main, R-based pipeline, which builds on R packages CRISPRseek³⁹, biomartr⁴⁰ and genbankr and can be either run from within an R environment, or called from the command line.

In addition, this pipeline registers for each found binding site a list of variables, including the number and position of mismatches between spacer and identified site, GC content of the sgRNA and if it contained undesired user-specified sequences such as user-defined bad seeds³⁴, and a corresponding estimate of the expected relative repression activity of the sgRNA on that site based on data from Hawkins et al.²¹ (default) or Qi et al.². If requested, the script also tests whether and to what extent the site falls within the boundaries on the nontemplate strand of an annotated genetic feature and the relative distance of the site to the start of the feature (for a gene the start codon). It can either or both output .csv files listing a summary of the binding sites with the first- and second-highest estimated repression activities per sgRNA and/or all sgRNA-site combinations and corresponding scores. We previously ran a former version of the pipeline for our D39V library for a range of commonly studied pneumococcal strains (D39V, R6, TIGR4, Hungary 19A-6, Taiwan 19F-14, 11A and G54; <https://veeninglab.com/crispri-seq>)¹¹. Downstream library efficiency analyses and the complete R pipeline and user manual are available on our GitHub page (<https://github.com/veeninglab/CRISPRi-seq>). This can be used to assess the expected efficiency of the readily available sgRNA library we constructed for the genome of interest, or to design a new sgRNA library (see below).

In the past, most sgRNA design rules and tools were optimized for eukaryotes and nuclease activity, or restricted to a selection of model organisms (e.g., <https://broad.io/crispick>)^{41–43}. Bacterial CRISPRi studies mostly use ‘in-house’ defined rules and scripts for sgRNA design, but these are often specific to the organism, experiments and growth conditions used, and they do not always allow for high-throughput (e.g., for all genes in a genome at once) or sparse (e.g., just one or a few sgRNA(s) per gene) library design^{7,15,21,34,44,45}. Here, we present an easy-to-customize pipeline that can design a sparse library of one or several optimal sgRNA(s) per annotated genetic feature for a whole local GenBank-formatted or National Center for Biotechnology Information-deposited genome in an automated fashion. First, all nontemplate strand-targeting candidate sgRNAs are identified for all annotated features. Second, all possible, relevant binding sites of each candidate on the genome are detected and scored, using the same method as described above. Third, the candidates are filtered, primarily based on expected off-target activity, to select one or more optimal sgRNA(s) per feature for which at least one zero-mismatch sgRNA exists (Supplementary Fig. 2). In addition, users can specify ‘bad seed’ or other sub-spacer sequences (such as specific restriction enzyme cut sites) they wish to exclude or avoid in sgRNA design.

Off-target activity is defined here as the maximum expected relative repression activity over all off-target binding sites per candidate sgRNA, as described before¹¹. Briefly, repression activity decreases as a function of the number and within-spacer position of mismatches between spacer and binding site^{2,3,21,33}. Repression activity of the sgRNA on a binding site is then estimated as the product of within-spacer mismatch-specific penalties, assuming the independence of mismatched base effects on repression activity² also holds for more than two and nonadjacent bases. This gives an expected activity on the [0,1] interval, which can be interpreted as a percentage of relative retained repression activity compared with a hypothetical zero-mismatch binding site at the same location. Custom mismatch-specific penalties can be passed as input parameter, but built-in penalties are provided from the *E. coli* and *B. subtilis* studies of Qi et al.² and Hawkins et al.²¹, with the default on nucleotide-median values for *B. subtilis* as reported by the latter. For each feature, we then calculate for each of its candidate sgRNAs the maximum off-target repression activity over all its respective binding sites. Following a minimax approach (i.e., taking the option with the best worst-case scenario), we subsequently select per feature the candidate sgRNA with the minimum maximum expected off-target effect, and those within a user-defined range of that minimum (Supplementary Fig. 2). Of this set of remaining candidates, the one with the shortest binding site distance to the start of the feature is picked as optimal, since smaller distances have been reported to correlate with increasing transcriptional repression². Note that this range parameter can be set to 0, placing all emphasis in optimal sgRNA selection on minimizing off-target activity without considering the distance of the sgRNA binding site to the start codon. This might be of interest for cases in which this distance was reported to have no effect on knockdown efficacy^{26,34}. Furthermore, the pipeline allows for any user-specified number of sgRNAs to be designed per feature, which might be useful if the user is concerned about the abovementioned downsides of a single sgRNA-per-feature setup.

In summary, our sgRNA design pipeline returns by default one, but optionally several sgRNA(s) per annotated feature, chosen with a primary selection on minimizing off-target effects and an optional secondary selection on maximizing proximity to the start of the genetic feature. Undesired sub-spacer sequences such as specific bad seeds or restriction enzyme cut sites can be excluded or avoided in the process, and mismatch-based repression activity estimation can be customized. It can also output the complete list of sgRNA candidates for each annotated feature plus all described scores, which allows the user to make their own sgRNA library selection, if the options offered by the input parameters of the pipeline are not satisfactory. In addition, oligos required for sgRNA construction are automatically designed and provided in the output as well, for which overhang sequences default to adhesive ends compatible with our cloning vector (Addgene #141090), but this can be easily customized as input parameter for different cloning strategies or cloning vectors (Fig. 3a). The produced list of sgRNAs can be further filtered by the user to pick one optimal sgRNA per operon, provided this information is available. Note, however, that although downstream polarity has been relatively well documented, reverse polarity effects appear to be species-dependent^{6–8,16,17,26,34}. The pipeline is very versatile, with standard input parameters that can easily be changed according to the user's experience or in the light of new data. Default parameters are set to values we found to reproduce >85% of our manually designed *S. pneumoniae* D39V sgRNA library. The pipeline can either be run from within an R environment or be called from the command line. The complete pipeline is available on GitHub, together with instructions for running it (<https://github.com/veeninglab/CRISPRi-seq>).

Vector design (Step 5)

To construct a well-balanced CRISPRi library, it is important to minimize potential bias that can be introduced by sgRNA library cloning. We thus adopted a PCR-free strategy for sgRNA cloning. The plasmid pPEPZ-sgRNAclone (Addgene #141090) was designed based on plasmid pPEPZ, an integration vector for *S. pneumoniae*, to enable cloning of sgRNAs by Golden Gate assembly (Fig. 3b)⁴⁶. Note that the pPEPZ-sgRNAclone vector replicates in *E. coli*, whereas it integrates into the chromosome in *S. pneumoniae*. There are three major considerations in vector design.

One consideration is to enable efficient and convenient sgRNA cloning for construction of the CRISPRi library. To this end, we implemented a simple red/white selection assay for visual identification of successful sgRNA cloning. The vector contains a 739 bp *mCherry* gene with BsmBI restriction sites on either side, at the future insertion site of the sgRNA base-pairing sequence (Fig. 3b). Thus, red colonies after plating indicate successful transformation of *E. coli* with pPEPZ-sgRNAclone. Subsequent correct insertion of the desired 20 bp sgRNA spacer replacing *mCherry* is marked by a white colony color after plating (Fig. 3c). Digestion of pPEPZ-sgRNAclone produces two bands in gel electrophoresis: the 4,086 bp vector backbone and the 739 bp *mCherry* fragment. So, the 739 bp DNA fragment can serve as an indicator of efficient BsmBI digestion. The vector backbone can be purified by gel extraction to eliminate the 739 bp *mCherry* fragment and the undigested plasmid. The clean and well-digested vector backbone ensures cloning with less false-positive colonies, caused by undigested vector or self-ligation. In practice, we observed no red colonies among the 70,000 collected colonies, demonstrating excellent efficiency.

The second consideration is to ensure that transcription of the sgRNA initiates efficiently and uniformly across the pool. Since the most predominant and efficient initiating nucleotide in *S. pneumoniae* is adenine¹⁹, we included an A as transcriptional start site (+1) right after the strong constitutive P3 synthetic promoter⁴⁷. The sgRNA sequence then immediately follows via the abovementioned Golden Gate cloning scheme. Note that the insertion of A as transcriptional start site leads to one extra adenine in the transcribed sgRNA on the PAM-distal side, but this does not influence sgRNA efficiency^{2,5}.

The last consideration is to facilitate Illumina sequencing later on, by already including elements of Illumina amplicons in the vector (Figs. 3 and 4). This allows for sgRNA quantification using just a one-step PCR for their amplification (Fig. 4).

Construction of the sgRNA plasmid pool in *E. coli* (Steps 2–40)

Targeting sequences were ordered from Synbio Technologies as two partially complementary 24 nt oligos (Fig. 3a). Annealing of the two oligos produced a DNA fragment with a 20 bp sgRNA base-pairing spacer sequence and a 4 nt overhang at each end. In this step, the annealing reaction was made individually for each sgRNA, providing access to individual sgRNAs for cloning single mutants for validation studies. The two 4 nt overhangs were designed to be compatible with the two adhesive

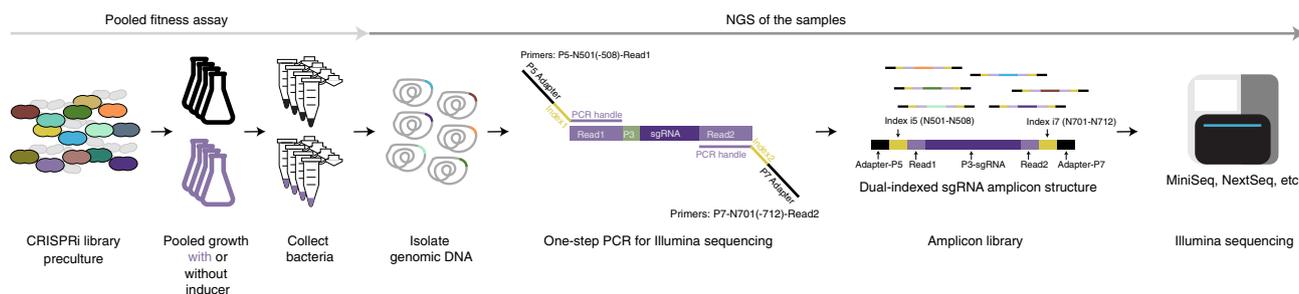


Fig. 4 | Workflow of the experimental CRISPRi-seq screening steps. The CRISPRi library is grown in a condition of interest either with (purple) or without (black) inducer. After sufficient generations of exponential growth, the bacteria are pelleted and their gDNA is isolated. The sgRNA sequences are extracted from the gDNA pool per sample by means of a one-step PCR. Primers are designed in such a way that they include the adapter sequences required for Illumina sequencing. Samples can be sequenced in a pooled fashion by barcoding the amplicons with sample-unique index combinations. NGS in this protocol was performed on an Illumina MiniSeq machine, with a custom recipe. Figure adapted from ref. ¹¹.

ends of the BsmBI-digested pPEPZ-sgRNA clone (Fig. 3a–c). Next, the annealed oligo pairs were pooled in equimolar amounts, followed by 5' phosphorylation (Fig. 3a). The fragments were then ligated into BsmBI-digested pPEPZ-sgRNA clone and used to transform chemically competent *E. coli* (Fig. 3b,c). Transformants were collected and cultured for plasmid purification (Fig. 3c). The purified plasmids serve as an sgRNA library reservoir for CRISPRi library construction in *S. pneumoniae*. The sgRNAs of the plasmid pool were sequenced by Illumina sequencing as a quality check for a well-balanced distribution of all sgRNAs (Fig. 3c). The pool of *E. coli* plasmids is available from Addgene (#170432).

The PCR-free, annealing-based approach we describe to clone sgRNAs is appropriate for the construction of compact libraries. In addition, with this approach, users have access to the individual sgRNA annealings to make single mutants for validation studies after the initial screening. However, for large sgRNA pools, e.g., in the case of CRISPRi libraries for bacteria with large genomes or with multiple sgRNAs per genetic feature, the annealing-based method is laborious. In those cases, a PCR-based method using oligo pools is more cost-efficient and practical. Several oligonucleotide manufacturers can provide ready-made pooled libraries, and such PCR-based approaches have been previously described^{6,34,48,49}.

Transformation of *S. pneumoniae* with pooled sgRNAs (Steps 46–61)

The host *S. pneumoniae* strain was induced with competence-stimulating peptide (CSP) for transformation with the sgRNA plasmid pool. In this protocol, we describe how to make the host strain with an IPTG-inducible *dcas9*. The transformants were collected and stored as CRISPRi library (Fig. 3d). In the CRISPRi library, both *dcas9* and the sgRNAs are integrated into the chromosome, and thus present with only one copy per cell.

Pooled library fitness assay (Steps 62–65)

The libraries were grown in C+Y medium treated with or without inducer, each in quadruplicate (Fig. 4). The induction treatment lasted for ~21 generations. In addition, we kept the bacteria in exponential growth phase during the treatment through three rounds of 1:100 culture dilutions into fresh C+Y medium, because of *S. pneumoniae* autolysis in stationary phase, again with or without inducer as appropriate (Figs. 2 and 4). Bacteria were collected after the treatment by centrifugation and stored at –80 °C for genomic DNA (gDNA) isolation. gDNA was isolated to serve as template for the following one-step PCR (Fig. 4). Determination of the duration of treatment should be based on screening analyses with samples collected at different timepoints. The optimal induction treatment aims to maximize the differences between the control and experimental conditions in order to capture more targets.

sgRNA sequencing with Illumina MiniSeq (Steps 66–73)

The amplicon libraries for Illumina sequencing can be prepared with a one-step PCR. The gDNA of the bacterial pellets was isolated for each sample and used as the PCR template. The sequences for Illumina index 1 and index 2 and the adapters were introduced by PCR (Fig. 4). Multiple samples can be pooled together with different combinations of index 1 and index 2 following the ‘Library pooling guidelines for the NextSeq and MiniSeq systems’ bulletin on the Illumina website.

Illumina sequencing platforms require diversified nucleotide sequences for effective template generation. Libraries generated by PCR amplification have little variation outside the sgRNA target region and are therefore challenging to sequence, resulting in lower yield and quality scores. The part of the PCR-produced amplicon to be sequenced here is allocated in between read 1 and read 2: the P3 promoter-driven sgRNA (P3-sgRNA, Fig. 4). It consists of the invariant 54 bp promoter sequence, followed by the varying 20 bp spacer sequence. To overcome the low-diversity issue of the amplicon library, we used a custom Illumina sequencing recipe that skips the common promoter sequence with 54 dark cycles. This turned out to be highly efficient. Demultiplexing was then carried out on the MiniSeq machine with local manager software (version 2.4.1), which returned one .fastq.gz file per sample. An alternative way is to do Illumina sequencing with a custom read 1 primer. One that worked well with the sgRNA amplicon library described in this protocol is 5'-CTTGA-CATTGCACTGTCCCCCTGGTATAATAACTATA-3'.

Based on our experiences and a naïve power analysis, we recommend using at least four replicates per treatment group, and an average of at least 250 reads per sgRNA, per sample, assuming that a \log_2FC of at least three can be reached for truly essential targets (Supplementary Fig. 3). So, if a MiniSeq High Output kit with 150 cycles and up to 25 million reads (Illumina, cat. no. FC-420-1002) is used, with our 1,499 sgRNA library, no more than $\frac{25,000,000}{(1,499 \cdot 250)} \approx 66$ samples should be loaded on a single run. If multiple sequencing runs are needed to cover all samples, be sure to distribute replicates over the runs so as to be able to detect and correct for any potential batch effects. If a pooled library, for instance, contains 10,000 unique sgRNA sequences, then such a MiniSeq run would only allow for roughly ten separate samples to maintain statistical power. Fortunately, if more sequencing depth is required, since Illumina libraries are universal, the same library can also be sequenced on other platforms such as the HiSeq (up to 400 million reads) and NovaSeq (up to 20 billion reads).

Data analysis (Steps 74–83)

Following sample sequencing and demultiplexing, the absolute abundance of each sgRNA per condition is obtained using the new 2FAST2Q tool presented here. 2FAST2Q was developed in Python to streamline all the initial sequencing data analyses into a single step. It automatically and easily performs read quality filtering, and sgRNA alignment for the input .fastq.gz files, returning a .csv file with all the absolute sgRNA read counts per condition for downstream analysis. 2FAST2Q is available as a compiled executable file, and thus requires no Python installations, being able to run as any other standalone program in MS Windows, macOS and Linux (Ubuntu). In addition, it is also available as PyPI module (<https://pypi.org/project/fast2q/>). All versions are available for download at Zenodo (<https://doi.org/10.5281/zenodo.5079789>). The original Python 3 code, user instructions and test data are also available on GitHub (<https://github.com/veeninglab/2FAST2Q>).

After parameter configuration, and upon initializing 2FAST2Q, the program unpacks all the available raw sequencing (.fastq.gz) files in the indicated folder. Reads are then trimmed to only obtain the nucleotides corresponding to the sgRNA positions (e.g., the first 20 bp of each read for the custom Illumina sequencing recipe described above). Simultaneously, a nucleotide-based quality filtering step is performed⁵⁰. A default Phred score of 30 is used, with all the trimmed reads with lower scoring nucleotides being discarded. The ensuing reads are then aligned to the sgRNAs based on a perfect match-only approach. Concomitantly, any remaining unmatched reads are then by default submitted to imperfect matching, where a one-nucleotide mismatch between sgRNA and sequenced read is allowed; this can, however, easily be adjusted to allow only for perfect matching by changing the input parameter, if desired. Reads for which the mismatch creates an alignment between more than one sgRNA are discarded. Finally, the absolute read count for each sgRNA is returned, both per sample and as a final compiled .csv file.

This compiled .csv file with raw counts is subsequently used as input for a differential enrichment analysis using the R package DESeq2 (ref. ⁵¹). For every sgRNA, its enrichment or depletion is assessed in the induced samples compared with the non-induced ones in terms of a \log_2FC , accompanied by a significance call using false discovery rate (FDR)-adjusted *P*-values. This is a proxy for the fitness of the targets corresponding to the sgRNAs (Fig. 5). The interaction term then quantifies differential gene fitness between two conditions. Other packages with similar goals are edgeR⁵² and limma⁵³, but when the design entails multiple sgRNAs targeting each gene, the use of MAGeCK is recommended⁵⁴. The latter scores gene fitness on the basis of fold changes of all sgRNAs targeting that gene, thereby decreasing the influence of individual sgRNA off-target effects⁵⁴. In addition, it allows for the inclusion of control sgRNAs that do not target any gene into the depletion analysis for normalization purposes, which might in general be wise to include regardless of the tool

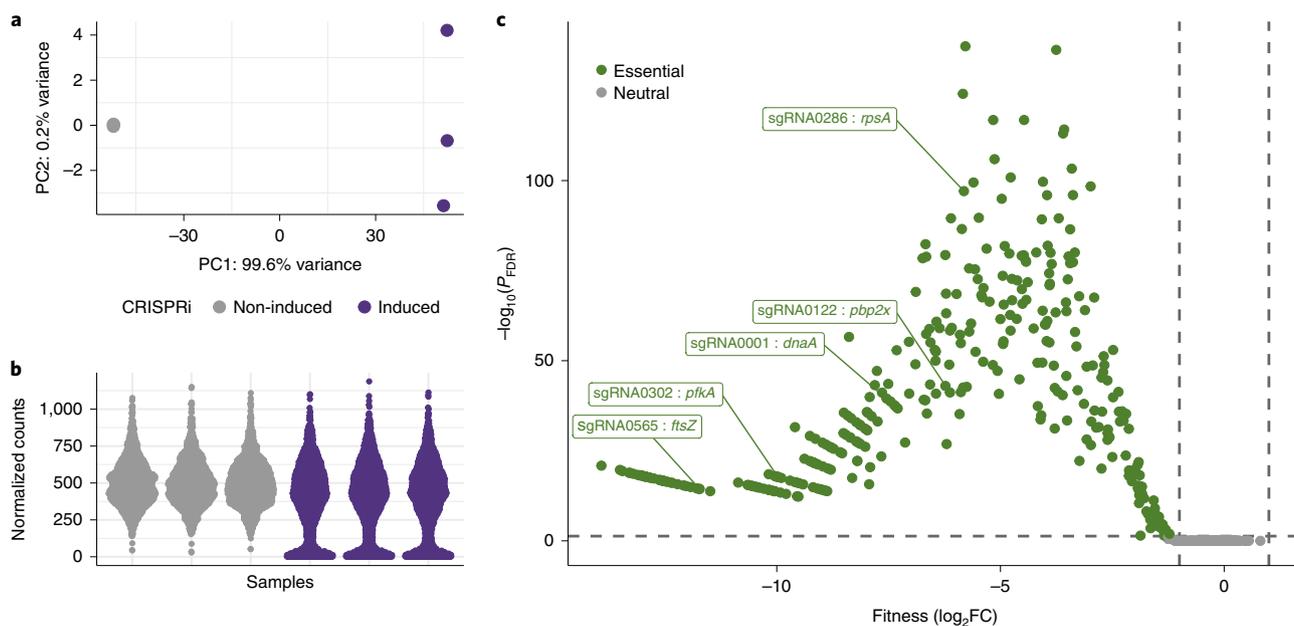


Fig. 5 | Example outcomes of differential sgRNA enrichment analyses. Published data of Liu et al.¹¹ were used to score gene fitness as a \log_2FC in its corresponding sgRNA abundance upon CRISPRi induction with 1 mM IPTG after 21 generations of exponential growth in C+Y medium. Cultures with and without inducer were both grown in biological triplicates. More analysis examples and explanations can be found in the DESeq2 vignette (<https://doi.org/10.18129/B9.bioc.DESeq2>). **a**, Principal component analysis shows similarity in sgRNA contents between samples, which is especially useful for experiments with multiple growth conditions. In general, the non-induced samples should be close together, even when cultured in different growth conditions, as there is no selection on sgRNAs in absence of dCas9. Here, all samples were grown in the same condition (C+Y) and almost all variance in sgRNA contents was explained by CRISPRi induction, with minimal variation among replicates. Similar information can be obtained by computation of pair-wise sample correlations and presenting these in heatmaps combined with hierarchical clustering. **b**, Normalized sgRNA counts show an approximately normal distribution of sgRNA abundance in non-induced samples, and an enrichment of low counts upon CRISPRi induction, brought about by sgRNAs targeting essential genes. Widths of the sample-wise scatter plots reflect local data density, as plotted with R package ggbeeswarm (<https://CRAN.R-project.org/package=ggbeeswarm>). **c**, Volcano plot of sgRNA target fitness scores in the in vitro C+Y laboratory condition. sgRNAs targeting genes known to be involved in several key processes in the cell are indeed significantly depleted (FDR-adjusted P -value < 0.05 , $\log_2FC < 1$; dotted gray lines) and thus marked essential: *dnaA* (DNA replication), *ftsZ* (cell division), *pfkA* (carbon metabolism), *rpsA* (translation) and *pbp2x* (cell wall biosynthesis). Output produced with DESeq2 (ref.⁵¹), normalized counts computed with its counts() function, and \log_2FC values shrunk with the ‘apeglm’ method⁵⁷.

being used, but in particular when most of the sgRNAs in the library are expected to be depleted (i.e., when most target genes are expected to be essential). As CRISPR screenings in general become more popular, these tools are even integrated into easy-to-use web applications⁵⁵.

If the differential enrichment analysis yields any specific sgRNA of interest, we recommend checking all potential binding sites of this sgRNA to ensure the observed effect is due to repression of the on-target gene it was designed for. Note, however, that it might be difficult to assign an observed phenotype to specific (off-)target sites, especially in the one-sgRNA-per-feature scenario; it is always important to validate any potentially responsible features with follow-up experiments. All such relevant binding sites can again be identified using the library efficiency evaluation pipeline described in the sgRNA design section. These tables are already available for up to six sgRNA–site mismatches in total for the pneumococcal strains we checked for our own sgRNA library on <https://veeninglab.com/crispri-seq>¹¹.

Materials

Biological materials

- *E. coli* competent cells stbl3 (Invitrogen cat. no. C737303)
- Transformable bacterium of interest. In the Procedure, we describe how to use the protocol on *S. pneumoniae* D39V (Public Health England, NCTC14078) **! CAUTION** *S. pneumoniae* is a biosafety level 2 organism and requires appropriate laboratory facilities, handling and disposal.

Reagents

- ddH₂O (Milli-Q (MQ) filtered water)
- 2× Phanta Max Master Mix (Vazyme, cat. no. P515)
- Esp3I (BsmBI; New England BioLabs, cat. no. R0734L)
- T4 DNA ligase (New England BioLabs, cat. no. M0202S)
- T4 Polynucleotide kinase (New England BioLabs, cat. no. M0201L)
- 10 mM ATP (New England BioLabs, cat. no. P0756S)
- GelRed nucleic acid stain (Biotium, cat. no. 41003)
- Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, cat. no. Q32851)
- Tween-20 (Sigma-Aldrich, cat. no. P9416-100ML)
- Spectinomycin (Applichem, cat. no. APA3834.0010)
- Tetracycline (Applichem, cat. no. APA2228.0025)
- Sodium dodecyl sulfate (Sigma-Aldrich, cat. no. L3771-25G)
- Sodium deoxycholate (Sigma-Aldrich, cat. no. D6750-10G)
- Ethanol absolute (BDH Chemicals, cat. 20821.321)
- Isopropanol (Fisher Chemical, cat. no. A416-500)
- LB broth, Difco (BD, cat. no. 244620)
- LB agar, Difco (BD, cat. no. 244520)
- Columbia blood agar base (OXOID, cat. no. CM0331)
- IPTG (Applichem, cat. no. APA1008.0025)
- Macherey-Nagel NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel, cat. no. 12303368)
- Plasmid Miniprep Kit, PureYield Plasmid miniprep system (Promega, cat. no. A1222)
- MiniSeq Reagent Kits
 - MiniSeq Mid Output Kit, 300 cycles (Illumina, cat. no. FC-420-1004)
 - MiniSeq High Output Kit, 150 cycles (Illumina, cat. no. FC-420-1002)
- PhiX Control Kit v3 (Illumina, cat. no. FC-110-3001)
- GeneRuler 1 kb DNA ladder (Thermo Scientific, cat. no. SM0311)
- 100 bp ladder (Invitrogen, cat. no. 15628019)
- Primers for the sgRNA pool (Synbio Technologies, synthesized into deep 96-well plates, dissolved with water to 100 μM)
- Primers for PCR (Integrated DNA Technologies, synthesized in individual tubes, dissolved with water to 100 μM)
- Sodium hydroxide (Sigma-Aldrich, cat. no. S8045-500G)
- Glycerol (Sigma-Aldrich, cat. no. G5516-1L)
- CSP type 1 (CSP-1; a 17 amino acid peptide; its sequence is EMRLSKFFRDFILQRKK; synthesized by GenScript)
- Sheep blood defibrinated (Oxoid Deutschland, cat. no. SR0051E)
- Tris-base (Sigma-Aldrich, cat. no. 648310-2.5KG)
- EDTA, disodium salt (Sigma-Aldrich, cat. no. 324503-1KG)
- DNA agarose (Eurobio Scientific, cat. no. GEPAGA07-65, 500g)

Plasmids

- pPEPZ-sgRNAclone (Addgene plasmid #141090)
- pPEPY-PF6-lacI (Addgene plasmid #85589)
- pJWV102-PL-dCas9 (Addgene plasmid #85588)
- sgRNA plasmid pool, if used (Addgene #170432)

Equipment

Hardware

- Qubit 4 fluorometer (Invitrogen, cat. no. Q33226)
- MiniSeq sequencing system (Illumina, cat. no. SY-420-1001)
- NanoDrop spectrophotometer (Thermo Scientific, NanoDrop One^C)
- Serological pipettes, 5, 10 and 25 ml (SARSTEDT, cat. nos. 86.1253.001, 86.1254.001 and 86.1685.001)
- Normal-size Petri dishes, 92 × 16 mm (SARSTEDT, cat. no. 82.1473.001)
- Square Petri dishes, 120 × 120 mm (Gosselin, cat. no. 030555S)
- Eight-strip PCR tubes, 0.2 ml, individually attached domed caps (Starlab, cat. no. I1402-2900)
- pH meter, FiveGo, pH/mV portable meter (Mettler Toledo, cat. no. MT30266882-1EA)

- 50 ml Falcon tubes (Greiner, cat. no. 227261)
- 15 ml Falcon tubes (Greiner, cat. no. 188261)
- 1.5 ml Eppendorf tubes, microtube (SARSTEDT, cat. no. 72.690.001)
- 2 ml Eppendorf tubes, microtube (SARSTEDT, cat. no. 72.695.500)
- Red-cap tubes, test tubes, sterile (VWR, cat. no. 212-1823)
- Inoculation loop, 10 μ l loop (COPAN, cat. no. 8177CSR40H)
- Cell spreader, L-shaped (COPAN, cat. no. 174CS01)
- Cryotubes, 1.2 ml (Nalgene, cat. no. 5000-0012)
- Glass beads, 3 mm (Huberlab, cat. no. 8.1401.03)
- 96-well plate (Corning Incorporated Costar, cat. no. 3595)
- Microplate reader (Tecan, Spark multimode microplate reader)
- Computer (although all the informatic analyses can be performed on a common personal computer, having at least 6 GB random-access memory is recommended and having eight or more logical processors speeds up the sgRNA design and evaluation scripts considerably)

Software

- MiniSeq Local Run Manager (version 2.4.2) for demultiplexing (<https://support.illumina.com/downloads/local-run-manager-for-miniseq-software-guide.html>)
- 2FAST2Q (accessible from GitHub <https://github.com/veeninglab/2FAST2Q>, PyPI <https://pypi.org/project/fast2q/>, Zenodo (<https://doi.org/10.5281/zenodo.5079789>).
- R version 4.0.2 or higher (<https://www.r-project.org/>)
- Python version 3.7 or higher (<https://www.python.org/downloads/>)
- Optional, but preferred: RStudio version 1.3.1073 or higher (<https://rstudio.com/>)
- sgRNA design and evaluation R scripts (<https://github.com/veeninglab/CRISPRi-seq>)
- R package DESeq2 (ref. ⁵¹) (<https://doi.org/doi:10.18129/B9.bioc.DESeq2>)
- R package CRISPRseek³⁹ (<https://doi.org/doi:10.18129/B9.bioc.CRISPRseek>)
- R package biomartr⁴⁰ (<https://cran.r-project.org/package=biomartr>)
- R package genbankr (<https://doi.org/doi:10.18129/B9.bioc.genbankr>)

Reagent setup

1 M Tris-Cl at pH 8.0

Dissolve 121.1 g Tris-base in 800 ml of MQ water. Stir the solution on a magnetic stirrer. Adjust the pH to 8.0 with HCl. Add MQ water to a final volume of 1 L. Autoclave and store at room temperature (RT, ~22 °C). The solution can be stored for up to 3 months.

0.5 M EDTA at pH 8.0

Add 186.1 g EDTA to 800 ml of MQ water. Stir the solution on a magnetic stirrer. Adjust the pH to 8.0 with NaOH. Note that the EDTA will not be dissolved until the pH of the solution is adjusted to ~8.0. Bring the final volume to 1 L with MQ water. Autoclave and store at RT. The solution can be stored for up to 3 months.

5 M NaCl

Add 58.44 g NaCl to 140 ml of MQ water. Stir the solution on a magnetic stirrer. Bring the final volume to 200 ml with MQ water. Autoclave and store at RT. The solution can be stored for up to 3 months.

TEN buffer (10 \times)

10 \times TEN buffer contains 100 mM Tris, 10 mM EDTA and 1 M NaCl, and the final pH is 8.0. Prepare the 10 \times TEN buffer according to the table below.

Component	Volume (ml)
1 M Tris-Cl at pH 8.0	10
0.5 M EDTA at pH 8.0	2
5 M NaCl	20
MQ water	68
<i>In total</i>	100

Store at RT for up to 3 months.

Sterile 80% (vol/vol) glycerol

Mix 80 ml glycerol with 20 ml MQ water, and autoclave at 121 °C for 20 min. Store at RT for up to 3 months.

70% ethanol (vol/vol)

Mix 70 ml of ethanol absolute with 30 ml of MQ water. Store at RT for up to 3 months.

LB plates

Dissolve LB agar powder in MQ water according to the instructions of the manufacturer. Autoclave at 121 °C for 20 min. Cool down to ~50 °C. Add antibiotics when necessary. Mix well and then pour 15 ml of the medium into one Petri dish (92 × 16 mm). Avoid bubbles while pouring to make a smooth surface on the agar plate. The plates can be stored at 4 °C for 1 month.

LB broth

Dissolve LB broth powder in MQ water according to the instructions of the manufacturer. Autoclave at 121 °C for 20 min. The broth can be stored at 4 °C for 1 month.

C+Y medium

This is a semi-synthetic medium; see Supplementary Methods for instructions to make it.

Columbia blood agar plates

Dissolve the Columbia agar powder in MQ water according to the instructions of the manufacturer. Autoclave at 121 °C for 20 min. Cool down to ~50 °C. Add 5% (vol/vol) sheep blood, and add antibiotics when necessary. Mix well and then pour into Petri dishes. For the square Petri dishes (120 × 120 mm), referred as big Columbia blood agar plates in this protocol, pour 30 ml of the medium into each plate; for the round Petri dishes (92 × 16 mm), referred to as small Columbia blood agar plates in this protocol, pour 15 ml of the medium into each plate. Avoid bubbles while pouring to make a smooth surface on the agar plate. The plates can be stored at 4 °C for 1 month.

Spectinomycin stock (100 mg/ml)

Weigh 1 g spectinomycin powder with an electronic scale, and dissolve it in 10 ml MQ water. Sterilize the solution by filtering with a 0.2 µm filter in a flow hood. Make aliquots in sterilized 1.5 ml Eppendorf tubes. Store the stocks at -20 °C; the solution can be kept for 1 year.

Tetracycline stock (5 mg/ml)

Weigh 0.1 g of tetracycline powder with an electronic scale, and dissolve it in 20 ml MQ water. Sterilize the solution by filtering with a 0.2 µm filter in a flow hood. Make aliquots in sterilized 1.5 ml Eppendorf tubes. Store the stocks at -20 °C; the solution can be kept for 1 year.

Gentamicin stock (40 mg/ml)

Weigh 1 g of tetracycline powder with an electronic scale, and dissolve it in 25 ml MQ water. Sterilize the solution by filtering with a 0.2 µm filter in a flow hood. Make aliquots in sterilized 1.5 ml Eppendorf tubes. Store the stocks at -20 °C; the solution can be kept for 1 year.

Equipment setup**MiniSeq setup**

Here we describe how to use a custom sequencing recipe with 54 dark cycles (adjust the number of dark cycles based on the position of sgRNA spacer on the amplicons as described above in the section 'sgRNA sequencing with Illumina MiniSeq'). The recipe can be obtained from Illumina technical support, and the recipe we used to generate the example results shown here is provided in Supplementary Software. However, we recommend contacting Illumina technical support before applying our recipe to make sure it suits your Illumina sequencing system.

The following manuals contain the required guidance:

- MiniSeq System Denature and Dilute Libraries Guide: <https://support.illumina.com/downloads/miniseq-denature-dilute-libraries-guide.html>
- MiniSeq System Guide: <https://support.illumina.com/downloads/miniseq-system-guide-100000002695.html>

- MiniSeq Local Run Manager Software Guide: <https://support.illumina.com/downloads/local-run-manager-for-miniseq-software-guide.html>
- Library pooling guidelines for the NextSeq and MiniSeq systems: <https://support.illumina.com/bulletins/2016/10/library-pooling-guidelines-for-the-nextseq-and-miniseq-systems.html>

Procedure

sgRNA design ● Timing 1 d

- 1 Either design a custom sgRNA library with option A, or check how well the ready-made sgRNA pool designed for *S. pneumoniae* D39V fits your genome of interest with option B. Required files and usage instructions can be found on our CRISPRi-seq GitHub page: <https://github.com/veeninglab/CRISPRi-seq>. Both pipelines have two versions: one that can be called from the command line (*_cmd.R) and one that can be run within an R environment (*.R). The former can be called with the Rscript command. For the latter, we recommend opening, editing and running the scripts in the RStudio integrated development environment.

(A) de novo sgRNA design per annotated feature

- (i) For the within-R version, open the file `sgRNA_library_design.R` and adapt the input settings in the first code section ('1. Settings') as desired. The other code sections should not be changed. For the command line version, input parameters are directly specified in the terminal; run

```
Rscript sgRNA_library_design_cmd.R --help
```

to see the input options. For either method, at least the variables `input_genome`, `outdir` and `TINDRidir` should be set. See the GitHub page for elaborate explanations on script usage, input and output. Note that the default parameters are optimized to reproduce our *S. pneumoniae* D39V sgRNA library and might not be suitable for the situation of interest.

- (ii) For the within-R version, run the whole script from within R(Studio) once all input parameters are set. For the command line version, run the script, possibly with additional options if desired, with the Rscript command. In its simplest form it will be of the following form:

```
Rscript sgRNA_library_design_cmd.R -g my_genome.gb -o /my/out-dir/path/ -t /my/TINDRidir/path/
```

where sgRNAs will be designed with default parameters for local GenBank-formatted genome `my_genome.gb`, output files will appear in the directory specified by `-o`, and the `TINDRi.py` file is saved in the folder directory specified by the `-t` flag. This step will likely take anywhere between 10 min (typical *S. pneumoniae*) and 3 h (*E. coli*), but this depends heavily on the size of the genome, number of chromosomes (i.e., plasmids), number of annotated features, set value of the `regions` and `max_mismatch_cum` parameters and, importantly, computer specifications. Also note that the pipeline might take longer the first time it is invoked, due to, e.g., package installations. By default, the pipeline will run with parallel processing, leaving only one logical processor unused. Note that requesting additional outputs `output_all_candidates`, but more so `output_full_list` and `detect_offtarget_genes_full`, might add substantial computation time to the pipeline run, the extent of which again depends on the aforementioned factors.

- (iii) (Optional) Select one sgRNA per operon, if operon structures are known. In this case, select for each operon the sgRNA that targets the first genetic element in the operon. Smaller sgRNA libraries are easier to handle, more cost-effective and better for avoiding potential bottleneck effects as, e.g., in in vivo models¹¹.
- (iv) Order the oligos corresponding to the selected sgRNA list in the output file `*_sgRNAs_optimal.csv`, from columns 'oligoForward' and 'oligoReverse'. They are automatically designed, based on user input, with as defaults complementary 20 nt spacer sequences and including 'TATA' and 'AAAC' overhangs for the forward and reverse primers, respectively, for the annealing in sgRNA generation (Fig. 3a). These overhangs are compatible with the adhesive ends of the BsmBI-digested pPEPZ-sgRNAclone (Fig. 3b,c) but can be easily adjusted by changing the input parameters. Note that the fourth nucleotide (A) on the forward primer corresponds to the TSS (+1) of the sgRNA.

(B) Efficiency analysis of existing sgRNA library on genome of interest

- (i) Download the files `sgRNA-library_spn-D39V.csv` and either `sgRNA_library_evaluation.R` (if running from within an R environment) or `sgRNA_library_evaluation_cmd.R` (if running from command line) from <https://github.com/veeninglab/CRISPRi-seq>. For the within-R version, open the script and adapt the input settings in the first code section ('1. Settings') as desired. The other code sections should not be changed. For the command line version, run

```
Rscript sgRNA_library_evaluation_cmd.R --help
```

in a terminal to see all input parameter options. For either version, at least the variables `input_genome`, `sgRNA_file`, `outdir` and `TINDRidir` should be set. See the GitHub page for elaborate explanations on script usage, input and output. Note that the default parameter values are mostly set to match our experimental setup and the user might want to adjust them, e.g., when using another organism. In case the user will evaluate another sgRNA library, please provide a corresponding `.csv` file following the same format as `sgRNA-library_spn-D39V.csv`.

- (ii) For the within-R version, run the whole script from within R(Studio) once all input parameters are set. For the command line version, run the script, possibly with additional options if desired, with the `Rscript` command. In its simplest form it will be of the following form:

```
Rscript sgRNA_library_evaluation_cmd.R -g my_genome.gb -s sgRNA-library_spn-D39V.csv -o /my/outdir/path/ -t /my/TINDRidir/path/
```

where sgRNAs specified in the file flagged by `-s` will be evaluated with default parameters for local GenBank-formatted genome `my_genome.gb`, output files will appear in the directory specified by `-o`, and the `TINDRi.py` file is located in the folder directory specified by the `-t` flag. This step will likely take in the order of minutes with our sgRNA library and default settings, but this depends on the size of the genome, number of chromosomes (i.e., plasmids), number of annotated features, number of sgRNAs in the library, set value of the `regions` and `max_mismatch_cum` parameters and computer specifications. Also note the pipeline might take longer the first time it is invoked, due to, e.g., package installations. By default, the pipeline will run with parallel processing, leaving only one logical processor unused. Note that requesting additional output `detect_off-target_genes_full` might add substantial computation time to the pipeline run, the extent of which again depends on the aforementioned factors.

- (iii) Perform downstream analyses of interest on the output.csv file(s). Examples are given in the file `Pneumococcal_sgRNA_library_efficiency_evaluation.pdf` and the corresponding Rmd source file on our GitHub, but note that they are based on the output of a previous version of the pipeline¹¹.
- (iv) If the sgRNA library covers the genome of interest sufficiently as judged by the user, continue to Step 41. Otherwise, design a new sgRNA library de novo, following the steps at Step 1A.

Construction of a custom sgRNA plasmid pool ● Timing 1-2 weeks

▲ **CRITICAL** Steps 2–40 need to be followed only if using a de novo library.

- 2 *Making the pool of annealed oligo pairs* (Fig. 3a). Anneal the oligo pairs in individual tubes by preparing per pair a mixture for the annealing reaction in eight-strip PCR tubes as shown in the table below.

Component	Amount per reaction (μl)	Alternative amount per reaction (μl)
Forward oligo (100 μM)	22.5	9
Reverse oligo (100 μM)	22.5	9
10× TEN buffer	5	2
In total	50	20

Mix well by vortexing. Collect the liquid at the bottom of the PCR tubes by brief centrifugation. Incubate at 96 °C for 5 min, and then slowly cool to RT. We recommend using a Thermocycler for the incubation using the following settings.

Steps	Denature	Gradually decreasing	Anneal
1	96 °C, 5 min		
2		96 °C to 25 °C, 0.1 °C/s	
3			25 °C, maintain

- 3 To make the equimolar annealing pool, transfer 5 µl of annealing mix from each tube into one 15 ml Falcon tube. Change tips between tubes to avoid cross-contamination. Note that, depending on the number of sgRNAs, you might need a larger or smaller container for the annealing pool. The rest of the annealing mixes in the individual tubes should be labeled well and stored at –20 °C for cloning of individual sgRNAs for potential follow-up validation experiments.

■ **PAUSE POINT** The annealing mixtures can be kept at –20 °C for years in convenience of individual sgRNA mutant construction.

▲ **CRITICAL STEP** To avoid pipetting bias as much as possible, we transfer 5 µl of annealing product from each tube into a tube for the annealing pool. We do not recommend using a smaller volume.

- 4 Set up the phosphorylation reaction using the components described in the table below.

Components	Amount (µl)
Annealing pool	2
10× Polynucleotide kinase buffer	2
10 mM ATP	2
T4 Polynucleotide kinase	1
ddH ₂ O	13
<i>In total</i>	20

Incubate the reaction at 37 °C for 30 min. Note that the commonly performed step of heat inactivation by incubating at 65 °C for 20 min is skipped, to avoid denaturation of annealed oligo pairs. After incubation, add 980 µl MQ water into the 20 µl of reaction mixture for a 100-fold dilution. The concentration of the annealed, phosphorylated primer pairs in the diluted reaction mixture is 45 nM.

- 5 *Ligating the targeting sequences into the plasmid backbone* (Fig. 3b). Order the pPEPZ-sgRNAclone vector from Addgene (#141090).

- 6 Transform the vector into competent *E. coli* stbl3 according to the manufacturer's instructions. The transformants should be incubated at 37 °C and selected on Luria-Bertani agar (LBA) plates with 100 µg/ml spectinomycin. The *E. coli* with pPEPZ-sgRNAclone then show up as bright red/pink colonies after incubation (~20 h) at 37 °C, which is a good indicator for positive transformants (Fig. 3b).

▲ **CRITICAL STEP** We recommend using *E. coli* stbl3 competence cells, but not other *E. coli* strains, as we found stbl3 can maintain the vector stably with high yield. However, in other *E. coli* strains, like DH5α, mutations on the vector were observed and the yield was lower. The commercial chemically stbl3 competent cells provide a transformation efficiency of >10⁸ CFU/µg plasmid. Alternatively, transformation by electroporation into *E. coli* might provide a higher transformation efficiency.

- 7 Pick up one positive colony from Step 6, restreak on a fresh spectinomycin-containing LBA plate, incubate overnight at 37 °C, pick a single pink colony to grow in 5 ml of LB broth containing 100 µg/ml spectinomycin and place at 37 °C in a shaking incubator (200 rpm). When the OD₆₀₀ reaches 0.6, make a stock by mixing 800 µl of bacterial culture with 200 µl of sterilized 80% (vol/vol) glycerol.

■ **PAUSE POINT** The stock can be stored at –80 °C for years.

- 8 Prepare the bacteria for plasmid isolation. Inoculate 100 µl of the *E. coli* stock in 20 ml of LB broth with 100 µg/ml spectinomycin. Incubate at 37 °C, 200 rpm overnight.
- 9 Collect the bacteria by centrifugation at 8,000g for 5 min in six 2 ml Eppendorf tubes with each 3 ml of culture. Then perform plasmid isolation with the PureYield Plasmid miniprep system according to instructions of the manufacturer. Quantify the product with a NanoDrop spectrophotometer.
- 10 Digest the pPEPZ-sgRNAclone with the restriction enzyme Esp3I (BsmBI), by setting up the reaction as shown in the table below. Incubate the reaction mixture at 37 °C for 3 h.

Component	Amount
pPEPZ-sgRNAclone plasmid	5 µg
Eps3I	4 µl
10× Cutsmart buffer	20 µl
ddH ₂ O	Up to 200 µl

▲ CRITICAL STEP We recommend incubation for 3 h to obtain a relatively low background of undigested vectors. Higher ratio of plasmid pPEPZ-sgRNAclone in the reaction than listed in the table is not recommended, to avoid richer residual of undigested vectors.

- 11 Load the digestion product into a 1.2% (wt/vol) DNA agarose gel along with a 1 kb marker. Two bands should show up: one lighter band at 739 bp representing the *mCherry* fragment, the other band at 4,086 bp representing the plasmid backbone.
- 12 Carefully purify the digested plasmid backbone band by gel extraction using the Macherey-Nagel kit according to the manufacturer's directions. Use a NanoDrop spectrophotometer to quantify the product.
- 13 Ligate the annealed oligo pool of Step 4 with the digested pPEPZ-sgRNAclone by setting up the reaction as described in the table below and incubating at 22 °C for 2 h or overnight. Note that the ligation is done using a molar ratio of 1:5 vector to insert.

Component	Amount	Final concentration
Digested pPEPZ-sgRNAclone	98 ng	3.6 nM
Diluted phosphorylated annealing pool of Step 4 ^a	4 µl	18 nM
T4 ligase	1 µl	-
10× T4 ligase buffer	1 µl	-
ddH ₂ O	Up to 10 µl	-

^aThe diluted phosphorylated annealing pool of Step 4 is 45 nM.

▲ CRITICAL STEP One tube of the reaction above can provide donor DNA for one transformation of 50 µl of the chemically competent *E. coli* stb13 strain. Multiple tubes of the same reaction might be required depending on transformation efficiency of the competent *E. coli* cells. Preliminary testing with the ligation product for transformation into *E. coli* is necessary to evaluate the number of tubes required (see Steps 14–30). Here, according to our experience, we recommend doing three tubes of ligation for three transformation experiments as described below. Users can adjust the number of tubes according to the transformation efficiency acquired.

- 14 *Transformation of ligated plasmid pool into E. coli stb13* (Fig. 3c). Set a water bath to 42 °C; warm up the LB medium and LBA plates with 100 µg/ml spectinomycin at 37 °C.
- 15 Thaw five vials of competent *E. coli* stb13 cells on ice for each transformation: one vial for a negative control with cut, but not ligated, vector from Step 12; one vial for a positive control with uncut vector; and the other three vials for transformation with the ligation product from Step 13.
- 16 Transfer 10 µl of ligation product from Step 13 into 50 µl of thawed, competent *E. coli* stb13 cells for each of the three vials of competent cells. Transfer 100 ng of cut, but not ligated, vector into the vial set as a negative control. Transfer 100 ng of uncut plasmid into the vial set as a positive control. Mix gently.
- 17 Incubate the tubes on ice for 30 min.
- 18 Heat-shock the tubes at 42 °C for 45 s, and then place the tubes on ice for 2 min.
- 19 Add 450 µl of prewarmed LB medium to each tube. Incubate the tubes at 37 °C, 200 rpm for 1 h.
- 20 After the 1 h recovery, pool the cells transformed with the ligation product together and mix well by inverting the tubes.
- 21 Prepare a dilution of the pooled cells transformed with the ligation product for calculation of the transformation efficiency. To do this, first add 10 µl of the cells into 990 µl of LB medium for a 100-fold dilution; then add 100 µl of the diluted cells into 900 µl of LB medium for a final 1,000-fold dilution. Plate 100 µl of both the 1,000-fold and 100-fold diluted cells on prewarmed LBA plates with 100 µg/ml spectinomycin. We recommend plating both dilutions to ensure countable plates are obtained, to be used in Step 25. Dilute and plate the positive transformation control in the same

- way. For the negative transformation control, plate 100 μl of the transformed cells without dilution onto LBA plates with 100 $\mu\text{g/ml}$ spectinomycin.
- 22 To plate the transformed cells for construction of the sgRNA pool, spread 100 μl of the transformants without dilution onto prewarmed LBA plates with 100 $\mu\text{g/ml}$ spectinomycin with glass beads.
 - 23 Incubate the plates at 37 °C overnight (~14–18 h).
 - 24 Check the color of colonies on the plates. More than 99% of the colonies should be white (Fig. 3c).
? TROUBLESHOOTING
 - 25 Check the plates with the negative and positive transformation controls. Few colonies (<50) should be present on the negative control plate, with many red colonies on the positive control plate.
? TROUBLESHOOTING
 - 26 Estimate the number of transformed colonies per sgRNA. Count the colonies on the plates that were prepared for calculations of the transformation efficiency with the ligation product (Step 21). Use this to calculate the total number of colonies on all plates, taking into account the dilution factor (100-fold or 1,000-fold) and the volume of pooled transformed cells. We recommend a coverage of at least 50 colonies per sgRNA.
▲ CRITICAL STEP It is critical to obtain a sufficient number of colonies to ensure that each sgRNA is adequately present in the sgRNA pool.
? TROUBLESHOOTING
 - 27 Collect the colonies from the LBA plates. Pipette 1 ml of LB medium with 100 $\mu\text{g/ml}$ spectinomycin onto the plate. Scrape the colonies off with a cell spreader, and transfer the liquid into a sterile 50 ml Falcon tube. Aim for at least 70,000 colonies for a ~50-fold theoretical coverage of the library.
 - 28 Repeat Step 27 for a second wash, to collect any remaining bacteria.
 - 29 Combine all the collected bacteria. Mix well by vortexing to make sure all the bacterial pellets are well dispersed. Measure the OD₆₀₀, and dilute the bacterial resuspension to OD₆₀₀ = 1.
 - 30 Add 0.25 \times sterilized 80% (vol/vol) glycerol into the OD₆₀₀ = 1 bacterial resuspension. For example, add 2.5 ml of 80% glycerol for 10 ml resuspension. Mix well, make 1 ml aliquots in cryotubes and store at –80 °C.
■ PAUSE POINT The aliquots of the pooled *E. coli* sgRNA library can be kept at –80 °C for at least a year.
 - 31 *Isolate the plasmid pool from E. coli stb13* (Fig. 3c). Thoroughly thaw one aliquot of the stock of *E. coli* pool from Step 30. Inoculate 500 μl into 20 ml of LB medium with 100 $\mu\text{g/ml}$ spectinomycin in a flask. Incubate at 37 °C, 200 rpm for 5–8 h.
▲ CRITICAL STEP It is critical to make sure the inoculum taken from the thawed stock contains a sufficient number of bacteria to have a good coverage of the sgRNA pool.
 - 32 Collect the bacteria by centrifugation at 8,000g for 5 min, and do a plasmid miniprep according to the instructions of the manufacturer.
 - 33 Pool the isolated plasmids, and quantify the concentration of plasmid using a NanoDrop spectrophotometer.
■ PAUSE POINT The plasmids containing the sgRNA pool can be aliquoted and stored at –20 °C for years.
 - 34 *Verifying sgRNA presence and distribution in the pool with next-generation sequencing (NGS)* (Fig. 3c). Confirm sgRNA spacer region diversity in the plasmid pool by Sanger sequencing with primer P5-F: 5'-AATGATACGGCGACCACC-3'. Then, perform a one-step PCR to prepare the library for NGS. Set up the PCR reaction to amplify the sgRNA target sequence with Illumina sequences, which include the adapter, barcode index and sequence for read primer binding (Fig. 4). The primers for the PCR reactions are listed in Table 1. The first part of each primer (before the bold sections of sequence) is the adapter sequence of P5 (forward primers) or P7 (reverse primers); sequences in bold are the index sequence, as N501–N508 in forward primers and N701–N712 in reverse primers. With dual indexing, these indexes provide $8 \times 12 = 96$ combinations, allowing for a pool of 96 different samples in one sequencing run.
▲ CRITICAL STEP To check the diversity of the sgRNA pool within the plasmid miniprep, we recommend using primer pairs P5-N501-read1 and P7-N701-read2 for an NGS run with only this one sample. However, to fulfill the capacity of the NGS, users can pool this sample with other samples by dual indexing for one run. For indexing strategies, refer to Step 36.
■ PAUSE POINT Dissolve all the NGS primers in MQ water to 100 μM to make stock solutions. Primers can be stored at –20 °C for years.
 - 35 Dilute the 100 μM primer stocks to 10 μM with MQ water.
 - 36 Assemble the PCR reaction on ice as shown in the table below.

Table 1 | Primers for Illumina sgRNA amplicon library preparation by the one-step PCR

Primer	5'-3' sequence	Purpose
P5-N501-read1	AATGATACGGCGACCACCGAGATCTACACT AGATCGCT CGTCGGCAGCGTCAGATGTGTATA	Custom forward primer
P5-N502-read1	AATGATACGGCGACCACCGAGATCTACAC CTCTCTAT TCGTCGGCAGCGTCAGATGTGTATA	Custom forward primer
P5-N503-read1	AATGATACGGCGACCACCGAGATCTACACT TATCTCTT TCGTCGGCAGCGTCAGATGTGTATA	Custom forward primer
P5-N504-read1	AATGATACGGCGACCACCGAGATCTACAC AGAGTAGAT CGTCGGCAGCGTCAGATGTGTATA	Custom forward primer
P5-N505-read1	AATGATACGGCGACCACCGAGATCTACAC GTAAGGAGT CGTCGGCAGCGTCAGATGTGTATA	Custom forward primer
P5-N506-read1	AATGATACGGCGACCACCGAGATCTACAC ACTGCATA TCGTCGGCAGCGTCAGATGTGTATA	Custom forward primer
P5-N507-read1	AATGATACGGCGACCACCGAGATCTACAC AAGGAGTAT CGTCGGCAGCGTCAGATGTGTATA	Custom forward primer
P5-N508-read1	AATGATACGGCGACCACCGAGATCTACAC CTAAGCCT TCGTCGGCAGCGTCAGATGTGTATA	Custom forward primer
P7-N701-read2	CAAGCAGAAGACGGCATAACGAGAT TCGCCTTA GTCTCGTGGGCTCGGAGATGTGTAT	Custom reverse primer
P7-N702-read2	CAAGCAGAAGACGGCATAACGAGAT CTAGTACGGT TCTCGTGGGCTCGGAGATGTGTAT	Custom reverse primer
P7-N703-read2	CAAGCAGAAGACGGCATAACGAGAT TTCTGCCT GTCTCGTGGGCTCGGAGATGTGTAT	Custom reverse primer
P7-N704-read2	CAAGCAGAAGACGGCATAACGAGAT GCTCAGGA GTCTCGTGGGCTCGGAGATGTGTAT	Custom reverse primer
P7-N705-read2	CAAGCAGAAGACGGCATAACGAGAT AGGAGTCCG TCTCGTGGGCTCGGAGATGTGTAT	Custom reverse primer
P7-N706-read2	CAAGCAGAAGACGGCATAACGAGAT CATGCCTA GTCTCGTGGGCTCGGAGATGTGTAT	Custom reverse primer
P7-N707-read2	CAAGCAGAAGACGGCATAACGAGAT GTAGAGAG GTCTCGTGGGCTCGGAGATGTGTAT	Custom reverse primer
P7-N708-read2	CAAGCAGAAGACGGCATAACGAGAT CCTCTCTGG TCTCGTGGGCTCGGAGATGTGTAT	Custom reverse primer
P7-N709-read2	CAAGCAGAAGACGGCATAACGAGAT AGCGTAGCG TCTCGTGGGCTCGGAGATGTGTAT	Custom reverse primer
P7-N710-read2	CAAGCAGAAGACGGCATAACGAGAT CAGCCTCGG TCTCGTGGGCTCGGAGATGTGTAT	Custom reverse primer
P7-N711-read2	CAAGCAGAAGACGGCATAACGAGAT TGCCTCTT GTCTCGTGGGCTCGGAGATGTGTAT	Custom reverse primer
P7-N712-read2	CAAGCAGAAGACGGCATAACGAGAT TCCTCTAC GTCTCGTGGGCTCGGAGATGTGTAT	Custom reverse primer

Bold letters represent the nucleotide sequence of the indexes.

Component	Amount per reaction (µl)	Final concentration
2× Phanta Max Master Mix	25 µl	1×
10 µM forward primer (P5-N5XX-read1)	2 µl	0.4 µM
10 µM reverse primer (P7-N7XX-read2)	2 µl	0.4 µM
Plasmid sgRNA pool as template	0.5 µg	10 ng/µl
MQ water	Up to 50 µl	

▲ CRITICAL STEP Using different forward and reverse primer pairs for dual indexing allows different libraries to be pooled together in a single Illumina run, which is more efficient and cost-effective. In addition, it is important to maintain color balance for each base of the index reads for Illumina sequencing. To choose the indexes for the samples that are pooled together in a single run, please follow the ‘Nextera Low Plex Pooling Guidelines’ provided by Illumina.

▲ CRITICAL STEP It is important to use a high-fidelity DNA polymerase for the PCR reaction here (in our case, Phanta), to minimize errors in the amplification.

37 Run the PCR with the following program.

Cycle number	Denature	Anneal	Extend
1	95 °C, 30 s ^a		
2-11	95 °C, 15 s	55 °C, 15 s	72 °C, 30 s
12			72 °C, 5 min

^aFor plasmids as template, such as this example, pre-denaturation takes 30 s, whereas for gDNA as template it takes 3 min.

▲ CRITICAL STEP With the amount of PCR template indicated in Step 36, 12 PCR cycles would produce enough amplicons. More cycles can be used to improve the yield when necessary. We confirmed that no detectable bias on sgRNA distribution was introduced by amplification with as many as 30 cycles¹¹. Note

that the PCR program might be different for a different DNA polymerase. We recommend following the instructions of the DNA polymerase manufacturer for the settings.

- 38 Load the PCR products into a 2% (wt/vol) agarose gel. The PCR should yield a 303 bp product. Perform gel extraction to purify the PCR products with the Macherey-Nagel NucleoSpin Gel and PCR Clean-up Kit according to the manufacturer's instructions. We recommend eluting the PCR products with MQ water. Alternatively, magnetic bead-based PCR purification systems, such as the ProNex Size-Selective Purification System of Promega, or the AMPure XP Cleanup and Size Selection system, can be used for purification of the PCR product.
- 39 Quantify the purified PCR products with the Qubit dsDNA HS Assay Kit, following the manufacturer's instructions.
- 40 Perform sequencing on an Illumina sequencer such as the MiniSeq according to the user manual. For more details about running the sequencing and data analysis, please refer to Steps 69–82. The required analysis here is to check the distribution of the sgRNA counts. If there is more than one sample, be sure to normalize the counts for library size, which can be done using the DESeq2 R function `counts(dds, normalized = TRUE)`. We anticipate these counts will approximately follow a normal distribution, meaning all sgRNAs are present in the pool in approximately equal amounts, with no sgRNAs missing from the pool. This can be checked with a beeswarm or violin plot (e.g., as for the sgRNA plasmid pool in Fig. 3c).

? TROUBLESHOOTING

Construction of a host strain with IPTG-inducible dCas9 ● Timing 1 week

▲ **CRITICAL** This section describes how to use the IPTG-inducible system to drive the expression of dCas9 in *S. pneumoniae* (Fig. 3d).

- 41 To make the IPTG-inducible promoter P_{lac} functional, the *lacI*-gmR fragment needs to be integrated into the chromosome of *S. pneumoniae* D39V. To start this, amplify the *lacI*-gmR fragment from plasmid pPEPY-PF6-*lacI* (ref. ⁵) with primers F-prs1/R-prs1:

F-prs1: 5'-GTGAAGTTATGAACATCATCGGTAAG-3'

R-prs1: 5'-GGTTCGCAAGCCATGGTTG-3'

▲ **CRITICAL STEP** These primers to amplify *lacI*-gmR from pPEPY-PF6-*lacI* can be acquired from Addgene (#85589).

- 42 Run 3 μ l of the PCR product on a 1% (wt/vol) agarose gel along with a 1 kb marker for electrophoresis analysis. The expected size of the amplified fragment is 2,760 bp.
- 43 Purify the fragment with the Macherey-Nagel NucleoSpin Gel and PCR Clean-up Kit according to the instructions of the manufacturer, and quantify the product with a NanoDrop spectrophotometer.
- 44 Transform the amplified fragment into *S. pneumoniae* D39V following the standard transformation protocol described in Box 1, and select the transformed cells with 40 μ g/ml gentamicin. This transformation yields an *S. pneumoniae* D39V strain with *lacI* (D39V, *prs1::lacI*).
- 45 Transform plasmid pJWV102-PL-dCas9 (ref. ⁵) (Addgene plasmid #85588) into this strain (D39V, *prs1::lacI*), following the standard transformation protocol (Box 1). Select the transformed cells with 0.5 μ g/ml tetracycline. This step constructs the *S. pneumoniae* D39V strain with an IPTG-inducible *dcas9* integrated on the chromosome.

■ **PAUSE POINT** The stock of the strain with an IPTG-inducible dCas9 can be stored at -80 °C for years.

Construction of a CRISPRi library by sgRNA pool transformation ● Timing 3 d

- 46 Transform the inducible CRISPRi host strain obtained at Step 45 with the sgRNA plasmid pool obtained at Step 33 (Fig. 3d). To do this, first make the stock and pre-competent cells (Tc) of the *S. pneumoniae* host strain as in the standard protocol (Box 1, steps 1–8). After this step, follow the adjusted transformation protocol steps as outlined below.

- 47 Thaw one tube of Tc (100 μ l). This can provide cells for ten transformation experiments.

- 48 Add 900 μ l of C+Y medium (Supplementary Methods) and 1 μ l of CSP-1 (100 μ g/ml). Mix well by inverting gently.

▲ **CRITICAL STEP** Note that we describe how to transform D39V here, which requires CSP-1. Different types of CSP might be needed for different strains. For example, TIGR4 uses CSP-2. There are six distinct CSPs identified in *S. pneumoniae*, and most of the strains use either CSP-1 or CSP-2 (ref. ⁵⁶).

- 49 Incubate in a water bath at 37 °C for 12 min to activate the competence system.

- 50 Prepare two 2 ml Eppendorf tubes. Add 200 ng of the sgRNA plasmid pool DNA to tube 1. Tube 2 will serve as control without donor DNA.

Box 1 | Transformation of *S. pneumoniae* D39V ● Timing 2 d

Procedure

Make stock culture

- 1 Inoculate one colony of *S. pneumoniae* into 2 ml of acid C+Y medium in a red-cap tube. Seal the cap of the tube, and incubate in a water bath at 37 °C.
- 2 When the OD600 reaches ~0.3, stop the growth on ice and make stock by mixing 800 µl of bacterial culture with 200 µl of sterilized 80% (vol/vol) glycerol.

■ **PAUSE POINT** Stock can be stored at –80 °C for at least 1 year.

Make pre-competent cells (Tc)

- 3 From the stock culture, inoculate 1:100 into acid C+Y medium, e.g., 30 µl of stock into 3 ml of acid C+Y medium, in a red-cap tube.
- 4 Incubate in a water bath at 37 °C until the OD600 reaches 0.11.
- 5 Stop the growth by incubating on ice for 5 min.
- 6 Centrifuge 1 ml of the culture at 8,000g for 2 min.
- 7 Remove 920 µl of supernatant, so that there is ~80 µl of supernatant left.
- 8 Resuspend the bacterial pellet with the 80 µl of supernatant, and add 20 µl of 80% (vol/vol) glycerol. Mix well, and label the tube as Tc.

■ **PAUSE POINT** The cells can be stored at –80 °C for months.

Transformation

- 9 Thaw one tube of Tc (100 µl), which can provide cells for ten transformation experiments.
- 10 Add 900 µl of C+Y medium and 1 µl of CSP-1 (100 µg/ml). Mix well by inverting gently.
- 11 Incubate in a water bath at 37 °C for 12 min to activate the Tc. The competence system of *S. pneumoniae* is induced in this step.

▲ **CRITICAL STEP** There are different types of CSP for *S. pneumoniae*. *S. pneumoniae* D39 responds to CSP-1. To transform different pneumococcal strains, choose the CSP type accordingly. CSPs are 17-amino-acid signal peptides. Most *S. pneumoniae* strains encode one of the allelic variants CSP-1 and CSP-2. The amino acid sequence of CSP-1 is NH₂-EMRLSKFFRDFILQRKK-COOH, and for CSP-2 it is NH₂-EMRISRIILDFLFRKK-COOH⁵⁸. There are six major types of CSPs identified for *S. pneumoniae*⁵⁶, and one can check the *comC* and *comD* genes of the strain of interest to identify its CSP type. CSP can be ordered as an unmodified synthetic peptide from several peptide synthesis companies.
- 12 Transfer 100 µl of the activated cells into a 1.5 ml Eppendorf tube containing the donor DNA for one transformation. The final concentration of donor DNA is 1 ng/µl. Always set up a control with no donor DNA. Mix the cells with DNA well by inverting gently.
- 13 Incubate in a water bath at 30 °C for 20 min.
- 14 Add 900 µl of fresh C+Y medium into each tube. Mix well by inverting gently. Then incubate in a water bath at 37 °C for 1.5 h.
- 15 Plate the transformation culture with different volumes onto Columbia sheep blood agar plates with appropriate selective antibiotics.
- 16 Incubate overnight at 37 °C, 5% CO₂.
- 17 Streak the colonies that show up on the plates onto a fresh Columbia sheep blood agar plate with the same antibiotic used for selection of the transformants. Incubate overnight at 37 °C, 5% CO₂.

▲ **CRITICAL STEP** *S. pneumoniae* forms chains in C+Y medium, and also in other laboratory media. During transformation, some cells in the chain can be transformed while other cells in the same chain may not be⁵⁹. In addition, collective resistance was observed for some antibiotics in *S. pneumoniae*⁶⁰. To avoid cheater cells, which are not transformed but can thus survive on the selective plate within the colonies, we recommend streaking the colonies on plates for a second selection.
- 18 Inoculate one colony into 2 ml of C+Y medium, and incubate in a water bath at 37 °C.
- 19 When the OD600 reaches 0.3, stop growth by incubating the tubes on ice.
- 20 Make stock of the culture by mixing 800 µl of bacterial culture with 200 µl of 80% (vol/vol) sterile glycerol.

■ **PAUSE POINT** The stock can be stored at –80 °C for years.

- 51 Transfer 200 µl of the activated Tc to each of the 2 ml Eppendorf tubes. Mix the cells with DNA well by inverting gently.
- 52 Incubate the tubes in a water bath at 30 °C for 20 min.
- 53 Add 1,800 µl of fresh C+Y medium to each tube. Mix well by inverting gently. Then incubate in a water bath at 37 °C for 1.5 h.
- 54 Prepare a dilution to calculate the transformation efficiency. Add 10 µl of the transformant culture to 990 µl C+Y medium and mix well for a 100-fold dilution. Then transfer 100 µl of the 100-fold dilution into a tube with 900 µl C+Y medium and mix well to get a 1,000-fold dilution.
- 55 Plate 300 µl of the 1,000-fold dilution onto a big Columbia sheep blood agar plate (square Petri dish, 120 × 120 mm) with 100 µg/ml spectinomycin using sterile glass beads. This plate is to count colonies to estimate the transformation efficiency in Step 57.
- 56 Spread the 2 ml original transformant culture obtained in Step 53 onto another six big Columbia sheep blood agar plates, 300 µl per plate, with 100 µg/ml spectinomycin using sterile glass beads. Discard the leftover culture. Incubate the plates at 37 °C, 5% CO₂ overnight (14–16 h).

▲ **CRITICAL STEP** It is important to plate equal volumes of transformant culture onto each plate to avoid growth differences of colonies among plates. Spread the transformed cells evenly on the plates as much as possible to avoid inter-colony competition. For this purpose, we suggest to use glass beads instead of cell spreaders.
- 57 Estimate the number of transformants. Count the number of colonies *C* on the 1,000-fold dilution plates obtained in Step 55. Calculate the total number of transformants *T* obtained from all six plates as $T = 6 \times 1,000 \times C$.

▲ CRITICAL STEP It is critical to obtain a sufficient number of transformants for an acceptable coverage of the sgRNA pool. In the *E. coli* transformation step, we aimed for at least a 50-fold theoretical coverage of each sgRNA in the pool. To prevent stochastic loss of sgRNAs, we recommend on average >100 pneumococcal transformants per sgRNA, i.e., >149,900 transformants for the 1,499 sgRNA pool. For pneumococcal strains with higher or lower transformation efficiency, the volume of pre-competent (Tc) cells can be adjusted accordingly prior to transformation.

58 *Harvest the transformants.* Pipette 2 ml of C+Y medium with 100 µg/ml spectinomycin onto each big Columbia sheep blood agar plate. Scrape the colonies off with a 10 µl inoculation loop, and transfer the liquid into a sterile 50 ml Falcon tube by pipetting.

59 Repeat Step 58 for a second wash, to collect any remaining bacteria.

60 Combine all the liquid. Harvest the cells by centrifugation (4 °C, 8,000g, 5 min). Resuspend with fresh C+Y medium with 100 µg/ml spectinomycin to OD600 = ~0.3–0.6. Mix the resuspended culture with 80% (vol/vol) glycerol in a 4:1 (vol/vol) culture to glycerol ratio as stock for long-term storage at –80 °C. The stock made here is the CRISPRi library.

61 *Make a pre-culture of the CRISPRi library.* Inoculate 100 µl of the stock into 4 ml of fresh C+Y medium with 100 µg/ml spectinomycin in a red-cap tube. Incubate in a water bath at 37 °C. When the OD600 reaches 0.3, stop growth by incubating on ice. Make stock of the pre-culture by mixing bacteria with 80% glycerol (vol/vol) in a 4:1 (vol/vol) culture-to-glycerol ratio.

▲ CRITICAL STEP An inoculum of the CRISPRi library stock should contain sufficient bacteria to achieve appropriate library coverage. We suggest at least 100 cells per sgRNA, so >149,900 bacteria for the sgRNA pool designed for *S. pneumoniae* D39V.

■ PAUSE POINT The CRISPRi library stock and pre-culture can be stored at –80 °C for at least 1 year.

Pooled CRISPRi fitness assay ● Timing 1 d

▲ CRITICAL We describe here an in vitro CRISPRi-seq screening in laboratory conditions, C+Y medium (Fig. 4). The screening in C+Y medium is performed over ~21 generations. Two groups, the control group and the experiment group, are set up in the screening, each with (at least) four biological replicates. The CRISPRi system in the control group is off due to absence of the inducer, whereas the CRISPRi system in the experiment group is on because the inducer is added. The aim of this screening is to refine the essential gene list of *S. pneumoniae*. A long duration time of the treatment of 21 generations is therefore used, in order to detect the sgRNAs targeting genes with a marginal influence on bacterial growth. As a result, the sgRNAs targeting genes that are critical for pneumococcal growth drop out from the population and the differences among these sgRNAs are not well determined. To achieve better resolution of CRISPRi-seq on all sgRNAs, multiple timepoints of sampling might be necessary.

62 Inoculate 40 µl of the pre-culture made in Step 61 into 4 ml of fresh C+Y medium with (experiment group) or without (control group) inducer, four replicates each, totaling eight growth cultures. 1 mM IPTG is added as inducer for the experiment group. To ensure >20 generations of induction and competition, the cultures are diluted 1:100 three times in fresh medium (see below). This step is already the first dilution. Incubate the cultures in a water bath at 37 °C.

63 When the OD600 reaches 0.3, transfer 40 µl of the culture into 4 ml of fresh C+Y medium with or without inducer, as appropriate. This is the second 1:100 dilution. Incubate the tubes in a water bath at 37 °C for further growth.

64 When the OD600 reaches 0.3 again, dilute 100 µl of the culture into 10 ml of fresh C+Y medium with or without inducer, respectively. This is the third 1:100 dilution. Use 10 ml instead of 4 ml of fresh C+Y medium in order to collect sufficient bacteria for gDNA isolation. Incubate the tubes in a water bath at 37 °C.

65 When the OD600 reaches 0.3, harvest the cells by centrifugation (8,000g, 5 min, 4 °C). Remove the supernatant carefully.

■ PAUSE POINT The bacterial pellets can be stored at –80 °C for months, to be subjected to gDNA isolation at a later time.

Next-generation sequencing ● Timing 3 d

▲ CRITICAL These steps describe how to use the Illumina MiniSeq machine (Fig. 4), but other Illumina sequencing platforms can be used instead.

66 *Isolation of gDNA.* Isolate the gDNA of each of the eight samples with the Wizard Genomic DNA Purification Kit (Promega), and quantify the product with a NanoDrop spectrophotometer. We recommend dissolving the gDNA in DNase-free water.

■ PAUSE POINT The gDNA can be stored at –20 °C for several months.

67 Prepare the library for Illumina sequencing with a one-step PCR. To do this, undertake Steps 34–37 to amplify the sgRNA with the gDNA as template. Use primers listed in Table 1 for dual indexing of the samples; this enables the sequencing of multiple samples in one Illumina sequencing run. Several indexing strategies can be adopted for the eight samples here; details are given in the ‘Library pooling guidelines for the NextSeq and MiniSeq systems’ bulletin on the Illumina website. For example, for the eight samples used as an example here, choose N701/N702/N703/N704 as index 1 and N501/N502 as index 2.

▲ CRITICAL STEP Note that gDNA is used as template here. To reduce the number of PCR cycles, we advise use of up to 4 µg of gDNA for each 50 µl of PCR mixture, which produces enough amplicons with as few as nine cycles. With less gDNA, more PCR cycles are needed. We have compared the amplicon libraries produced by PCR with 10 cycles, 20 cycles and 30 cycles from the same input gDNA, and no substantial difference was observed.

68 Purify and quantify the amplified NGS library as described in Steps 38–39. This quantifies the purified amplicons in ng/µl.

69 *Prepare the library for sequencing.* Calculate the molar concentrations of the purified amplicons by converting the concentration in ng/µl to nM using the following formula:

$$\frac{\text{Concentration in ng/}\mu\text{l}}{660 \frac{\text{g}}{\text{mol}} \times \text{Amplicon size in bp (here 303)}} \times 10^6 = \text{Concentration in nM.}$$

70 Dilute and denature the library according to the Illumina standard protocol ‘miniseq-denature-dilute-guide-100000002697’; follow ‘Protocol A: Standard normalization method’. We advise using a 20% PhiX control. Keep the ready-to-load library on ice while preparing the sequencing cartridge.

71 Choose a MiniSeq kit. To sequence the sgRNA library here, we recommend using the MiniSeq Mid Output Kit, or MiniSeq High Output Kit (150 cycles) depending on the number of samples and aimed sequencing depth. For single-end sequencing, the Mid Output Kit yields 7–8 million reads and the High Output Kit yields 22–25 million reads. We recommend an average sequencing depth of at least 250 reads per sgRNA, per sample, based on experience and a naïve simulation study (Supplementary Fig. 3). For eight samples and 1,499 sgRNAs, $8 \times 1,499 \times 250 = 2,998,000$ reads are needed and, thus, the Mid Output Kit suffices.

72 Sequence the library on the Illumina MiniSeq according to the manual ‘miniseq-system-guide-100000002695’ provided by the manufacturer. Run the machine in a Standalone mode to use the custom sequencing recipe, which can either be obtained from the Illumina technical support or found in the Supplementary Software, for both Mid and High Output kits. If users decide to use the custom recipe we provided here, we suggest to confirm this choice with Illumina technical support as there might be differences between types of sequencing machines. The custom sequencing recipe is designed to avoid the issues caused by low-diversity sequences of amplicons. The first 54 cycles in the run are for sequencing of common sequences of the amplicons, and are thus set as dark cycles (i.e., the chemical reactions still take place, but no imaging). The following 20 cycles are to sequence the diversified base-pairing region of sgRNAs. Key parameters to set up the MiniSeq are listed in the table below.

Option	Parameter
Custom recipe	Mid-54dk (for Mid Output Kit) or High-54dk (for High Output Kit) ^a
System configuration	Standalone mode
Index 1	8 bp
Index 2	8 bp
Read length	97 bp ^b
Single or paired end	Single-end

^aThese are the custom sequencing recipes provided by Illumina technical support and in the Supplementary Software. ^bRead length ≥ 26 bp is sufficient to get the sgRNA sequence. Please note that the MiniSeq platform requires at least 26 cycles to estimate the ‘clusters passing filter’. For the results presented here, we set the read length to 97 bp, so the total number of chemistry cycles of this single-end sequencing run is 151 (97 + 54), which exploits the maximum capacity of the MiniSeq High Output Kit (150 cycles). However, as few as 26 cycles are sufficient to get the sgRNA sequences here.

- 73 Generate .fastq.gz files by requeuing the raw data with the Local Run Manager (version 2.4.1). MiniSeq runs with the custom recipe produce raw sequencing data without performing demultiplexing. To generate the (compressed) .fastq files for each sample, requeue the raw data by using the Local Run Manager following the guidelines of Illumina. Refer to ‘Local Run Manager Software Guide 1000000002702’ for how to ‘Requeue Analysis from Results Page’. Store the generated files on a network-attached storage server or an external drive, and preferably also make a back-up.
- **PAUSE POINT** The produced .fastq.gz files can directly be used as input for 2FAST2Q to obtain the raw counts per sgRNA per sample, as described below.

sgRNA counting using 2FAST2Q ● Timing 2 h

- 74 Obtain a .csv file with the nucleotide sequences of all used sgRNAs, and their respective names. See the provided D39V_guides.csv sample file on GitHub or PyPI (<https://pypi.org/project/fast2q/>).
- 75 If using either the MS Windows or the macOS executable 2FAST2Q program, download the appropriate software version (can be obtained from GitHub at <https://github.com/veeninglab/2FAST2Q>, or directly from Zenodo at <https://doi.org/10.5281/zenodo.5079789>), and proceed to Step 80. If using the non-executable version, proceed to Step 76 for parameter configuration.
- 76 Install the Python 3 package 2FAST2Q by typing the following code into a command window on a system with both Python 3 and pip installed:

```
pip install fast2q
```

- 77 *Configure the input.*

To open 2FAST2Q using the graphical interface method, type:

```
python -m fast2q
```

or, opening 2FAST2Q in command window mode, type:

```
python -m fast2q -c
```

- 78 (Optional) If using uncompressed .fastq files, change the ‘Input sequencing file extension’ option (--SE on the command line). At the moment, only .fastq or .fastq-like files are accepted.
- ▲ **CRITICAL STEP** All the raw sequencing files present in the indicated folder will be submitted to analysis. Make sure only the correct files are in the selected folder.
- 79 (Optional) Change the optional parameters section to match the experimental setup (an in-depth user guide is available at <https://pypi.org/project/fast2q/> or <https://github.com/veeninglab/2FAST2Q>).
- 80 *Run 2FAST2Q.* For the MS Windows and macOS versions: double-click the 2FAST2Q icon, and follow the on-screen configuration prompts (see Steps 78–79). The program should run automatically without any further user input. Exit the program when the ‘Analysis successfully completed’ message appears, or when the ‘compiled.csv’ file appears in the output folder.

? TROUBLESHOOTING

- 81 Upon completion, and depending on the user input, several files should be present in the indicated output folder: (1) the uncompressed .fastq files, (2) *_reads.csv files corresponding to the read counts per sgRNA per inputted sequencing file, (3) a compiled_stats.txt file containing all the relevant input/output information about the 2FAST2Q analysis, (4) a plot with the number of quality passed reads, and (5) a compiled.csv file with a compilation of all the read counts per sgRNA in all the input files. Use this latter file for all downstream analyses.

Differential fitness analysis ● Timing 1 d

▲ **CRITICAL** Testing for significant differential enrichment of sgRNAs in induced versus non-induced samples can be done with existing tools, e.g., the here used R package DESeq2 (ref. ⁵¹). If you have designed multiple sgRNAs per gene, it is better to use MAGeCK instead of DESeq2 (ref. ⁵⁴).

- 82 Follow the instructions of the DESeq2 vignette, which can be found at <https://doi.org/doi:10.18129/B9.bioc.DESeq2>. For fitness values within a condition, contrast the CRISPRi induction effect within that condition (i.e., the main effect). For differential fitness values between two conditions, take the interaction effect (i.e., the difference in CRISPRi induction effect between two conditions). See Fig. 5 for examples of

processed DESeq2 output of our IPTG-inducible *S. pneumoniae* D39V CRISPRi library cultured in exponential growth phase for 21 generations in C+Y medium¹¹, as described in this protocol. We advise using shrunk log₂FC values when displaying or using fitness scores for follow-up analyses⁵¹.

- 83 If any sgRNA hits are obtained that warrant follow-up experiments, check all possible binding sites for the sgRNA that could have brought about the observed phenotype by running the library efficiency pipeline as described in Step 1B, if this was not done yet.

(Optional) Follow-up validation experiments ● Timing 1–2 weeks

- 84 Clone the individual sgRNAs into the pPEPZ-sgRNAclone backbone using the annealing pairs for each sgRNA produced at Step 3. Ligate the annealing pairs of individual sgRNAs into the pPEPZ-sgRNAclone as described for the sgRNA pool (in Steps 4–13).
- 85 Transform the ligation product of the sgRNAs and digested pPEPZ-sgRNAclone directly into the host *S. pneumoniae* strain constructed at Steps 41–45, following the standard transformation protocol (Box 1). All the 10 µl of ligation product is added to 100 µl of activated Tc as donor DNA. We recommend plating 100 µl of the transformant culture onto Columbia sheep blood agar with 100 µg/ml spectinomycin. In addition, to capture well-separated colonies of the transformants, plate different dilutions of the transformant culture, including 100-fold and 10-fold dilutions.
- 86 *Validation of the sgRNA cloning.* Carry out a PCR to amplify the sgRNA for Sanger sequencing. First prepare the reaction according to the following setup.

Component	Amount (µl)
2× Phanta Max Master Mix	25
Primer P5-F (10 µM)	2
Primer P7-R (10 µM)	2
Stock of the bacterial culture	2
MQ water	19
<i>In total</i>	50

P5-F: 5'-AATGATACGGCGACCACC-3'

P7-R: 5'-CAAGCAGAAGACGGGCATACGAG-3'

Run the PCR with the following program.

Cycle number	Denature	Anneal	Extend
1	95 °C, 7 min ^a		
2–30	95 °C, 15 s	55 °C, 15 s	72 °C, 30 s
31			72 °C, 5 min

^aSince a bacterial culture is used as template in the PCR, we recommend extending the denaturation cycle to 7 min to break down the bacteria by heating.

- 87 Verify the cloning of the correct sgRNA by Sanger sequencing with P5-F. Align the sequencing results with the pPEPZ-sgRNAclone sequence to ensure the 20 nt base-pairing sequence of the sgRNA is inserted between the P3 promoter and the sgRNA scaffold.
- 88 *Verify the phenotype by applying the screening condition to the validation sgRNA mutant.* First make a pre-culture of the sgRNA mutant as described in Step 61.
- ▲ CRITICAL STEP** These steps describe how to screen in C+Y medium for evaluation of gene fitness, as an example that shows the setup for a validation study.
- 89 Dilute the pre-culture of the sgRNA mutant 1:100 into C+Y medium with (induced group) or without (non-induced group) 1 mM IPTG, with three replicates per group.
- 90 Aliquot 300 µl of each diluted culture into a well of the 96-well plate.
- 91 Set up the growth curve assay by measuring OD600 every 10 min for 18 h at 37 °C with a Tecan microplate reader. Please adjust the volume of diluted culture added into the 96-well plate according to the plate and the microplate reader used.
- 92 Plot the growth curves (e.g., with R or Prism GraphPad). Typical growth curves for sgRNAs targeting essential and dispensable genes are shown in Fig. 1.

Troubleshooting

Troubleshooting advice can be found in Table 2.

Step	Problem	Possible reason	Solution
24	There are >1% red colonies on the plates	Residual undigested pPEPZ-sgRNAclone	Increase the amount of Eps3I in the digestion of Step 10 Decrease the input of pPEPZ-sgRNAclone in the reaction of Step 10 Extend the incubation time of Step 10 Cut a sharp band for gel extraction of the digested backbone at Step 12
25	There are lots of colonies on the negative plate or few colonies on the positive plate	Residual undigested pPEPZ-sgRNAclone <i>E. coli</i> stbI3 competent cells are not efficient	Improve the digestion of pPEPZ-sgRNAclone with the same solutions for troubleshooting Step 24 Use efficient <i>E. coli</i> stbI3 competent cells
26	Insufficient number of colonies obtained by the transformation	<i>E. coli</i> competent cells are not efficient Ligation reaction is not efficient	Use more efficient <i>E. coli</i> competent cells in Step 15. Alternatively, try transformation by electroporation, which might provide a higher transformation efficiency Optimize Step 13 to improve the ligation efficiency by tuning the molar ratio of backbone and annealing pool Increase the number of ligation reactions and transformants
40	sgRNA counts are not approximately normally distributed	If multiple samples are compared in one plot: sgRNA counts not normalized Failure of sgRNA pool construction	Normalize the counts for library size, e.g., with the counts() function in DESeq2 This rarely happens if all the earlier steps are performed well. If it happens, re-do the pooling of the annealing mixture with careful pipetting to ensure equimolarity of each annealing in the solution at Step 3
80	Error message: Warning!! X and Y share the same sequence. Only X will be considered valid Error message: Check the path to the sgRNA file. No file found in the following path: X Error message: Please specify --s,--g,--o, and --se parameters. type -h into cmd for help Error message: Check the path to the X files folder. No files of this type found Program does not initialize Program crashes, or behaves unexpectedly	X and Y correspond to sgRNA names. The indicated entries have the same sequence, and only the first will be considered valid Wrong sgRNA file path specification in the input .txt file Not all parameters were inputted properly. Only happens in command window mode Wrong sequencing file directory path specification Wrong version of program installed: incompatible with user operating system	Not an issue; continue as planned. If unexpected, confirm the indicated sgRNAs have identical sequences Confirm the indicated path (X) to the sgRNA .csv file is correct Please follow the instructions to correct the input Confirm the indicated path to the folder with the sequencing (X) files is correct Confirm the downloaded program is the appropriate one for the current operating system. Contact the 2FAST2Q developer if the issue persists Check if the program behaves as expected with the provided sample data. Contact the 2FAST2Q developer if the issue persists

Timing

Times indicated below are representative for the examples described here; the procedure may take longer for different organisms, sgRNA library sizes, cloning techniques, growth conditions, number of samples, sequencing methods and computers.

Step 1, sgRNA design: 1 d

Steps 2–40, construction of a custom sgRNA plasmid pool: 1–2 weeks

Step 41–45, construction of a host strain with IPTG-inducible dCas9: 1 week

Steps 46–61, construction of a CRISPRi library by sgRNA pool transformation: 3 d

Steps 62–65, pooled CRISPRi fitness assay: 1 d

Steps 66–73, NGS: 3 d
Steps 74–81, sgRNA counting using 2FAST2Q: 2 h
Step 82–83, differential fitness analysis: 1 d
Steps 84–92, (optional) follow-up validation experiments: 1–2 weeks

Anticipated results

Completing Step 1 will yield either a set of oligos that can be used in Steps 2–40 to construct a pool of sgRNAs to cover the whole genome of interest, or an estimate of how well that genome is covered by our library, which can then be ordered from Addgene (#170432) for further use in the procedure as of Step 46. Steps 41–45 result in a suitable CRISPRi host strain, in which the library can be integrated according to Steps 46–61. Subsequently, pooled assays (Steps 62–65) give barcoded gDNA to be sequenced in Steps 66–73, which results in one .fastq.gz per sample. Applying 2FAST2Q (Steps 74–81) to these files yields raw read counts per sgRNA in a .csv file. Ultimately, differential enrichment analysis according to Steps 82–83 returns fitness scores per sgRNA and their corresponding target(s), and significantly enrichment or depletion calls, signaling costly or essential target genes for the condition being studied (see Fig. 5 for examples of results we have obtained). Optionally, following Steps 84–92 results in individual CRISPRi strains for follow-up studies of any genes identified that are of interest (for example, Fig. 1 shows a typical growth curve seen for an individual CRISPRi strain targeting an essential gene).

Data availability

Source data are provided with this paper.

Code availability

Source code for the sgRNA design and evaluation pipelines, as well as for the naïve power analysis and corresponding data, can be found at <https://github.com/veeninglab/CRISPRi-seq>. 2FAST2Q code, files and executables can be found at <https://github.com/veeninglab/2FAST2Q>, <https://pypi.org/project/fast2q/> or <https://doi.org/10.5281/zenodo.5079789>. (Off-)target binding site tables from Liu et al.¹¹ can be found through our website: <https://www.veeninglab.com/crispri-seq>.

References

1. Cain, A. K. et al. A decade of advances in transposon-insertion sequencing. *Nat. Rev. Genet.* **21**, 526–540 (2020).
2. Qi, L. S. et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).
3. Bikard, D. et al. Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res.* **41**, 7429–7437 (2013).
4. Lee, H. H. et al. Functional genomics of the rapidly replicating bacterium *Vibrio natriegens* by CRISPRi. *Nat. Microbiol.* **4**, 1105–1113 (2019).
5. Liu, X. et al. High-throughput CRISPRi phenotyping identifies new essential genes in *Streptococcus pneumoniae*. *Mol. Syst. Biol.* **13**, 931 (2017).
6. Wang, T. et al. Pooled CRISPR interference screening enables genome-scale functional genomics study in bacteria with superior performance. *Nat. Commun.* **9**, 2475 (2018).
7. Peters, J. M. et al. A comprehensive, CRISPR-based functional analysis of essential genes in bacteria. *Cell* **165**, 1493–1506 (2016).
8. Zhao, C., Shu, X. & Sun, B. Construction of a gene knockdown system based on catalytically inactive (“dead”) Cas9 (dCas9) in *Staphylococcus aureus*. *Appl. Environ. Microbiol.* **83**, e00291–17 (2017).
9. Larson, M. H. et al. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat. Protoc.* **8**, 2180–2196 (2013).
10. de Wet, T. J., Winkler, K. R., Mhlanga, M., Mizrahi, V. & Warner, D. F. Arrayed CRISPRi and quantitative imaging describe the morphotypic landscape of essential mycobacterial genes. *eLife* **9**, e60083 (2020).
11. Liu, X. et al. Exploration of bacterial bottlenecks and *Streptococcus pneumoniae* pathogenesis by CRISPRi-Seq. *Cell Host Microbe* **29**, 107–120.e6 (2021).
12. de Wet, T. J., Gobe, I., Mhlanga, M. M. & Warner, D. F. CRISPRi-Seq for the identification and characterisation of essential mycobacterial genes and transcriptional units. Preprint at *bioRxiv* <https://doi.org/10.1101/358275> (2018).
13. Beuter, D. et al. Selective enrichment of slow-growing bacteria in a metabolism-wide CRISPRi library with a TIMER protein. *ACS Synth. Biol.* **7**, 2775–2782 (2018).
14. Yao, L. et al. Pooled CRISPRi screening of the cyanobacterium *Synechocystis* sp PCC 6803 for enhanced industrial phenotypes. *Nat. Commun.* **11**, 1666 (2020).

15. Calvo-Villamañán, A. et al. On-target activity predictions enable improved CRISPR-dCas9 screens in bacteria. *Nucleic Acids Res* **48**, e64 (2020).
16. Rousset, F. et al. Genome-wide CRISPR-dCas9 screens in *E. coli* identify essential genes and phage host factors. *PLOS Genet* **14**, e1007749 (2018).
17. Vigouroux, A. & Bikard, D. CRISPR tools to control gene expression in bacteria. *Microbiol. Mol. Biol. Rev.* **84**, e00077–19 (2020).
18. Rousset, F. & Bikard, D. CRISPR screens in the era of microbiomes. *Curr. Opin. Microbiol.* **57**, 70–77 (2020).
19. Slager, J., Aprianto, R. & Veening, J.-W. Deep genome annotation of the opportunistic human pathogen *Streptococcus pneumoniae* D39. *Nucleic Acids Res.* **46**, 9971–9989 (2018).
20. Donati, S. et al. Multi-omics analysis of CRISPRi-knockdowns identifies mechanisms that buffer decreases of enzymes in *E. coli* metabolism. *Cell Syst.* **12**, 56–67.e6 (2021).
21. Hawkins, J. S. et al. Mismatch-CRISPRi reveals the co-varying expression–fitness relationships of essential genes in *Escherichia coli* and *Bacillus subtilis*. *Cell Syst.* **11**, 523–535.e9 (2020).
22. Jiang, W., Oikonomou, P. & Tavazoie, S. Comprehensive genome-wide perturbations via CRISPR adaptation reveal complex genetics of antibiotic sensitivity. *Cell* **180**, 1002–1017.e31 (2020).
23. Bosch, B. et al. Genome-wide gene expression tuning reveals diverse vulnerabilities of *M. tuberculosis*. *Cell* **184**, 4579–4592.e24 (2021).
24. Momen-Roknabadi, A., Oikonomou, P., Zegans, M. & Tavazoie, S. An inducible CRISPR interference library for genetic interrogation of *Saccharomyces cerevisiae* biology. *Commun. Biol.* **3**, 723 (2020).
25. Du, D. et al. Genetic interaction mapping in mammalian cells using CRISPR interference. *Nat. Methods* **14**, 577–580 (2017).
26. Rock, J. M. et al. Programmable transcriptional repression in mycobacteria using an orthogonal CRISPR interference platform. *Nat. Microbiol.* **2**, 16274 (2017).
27. Guzzo, M., Castro, L. K., Reisch, C. R., Guo, M. S. & Laub, M. T. A CRISPR interference system for efficient and rapid gene knockdown in *Caulobacter crescentus*. *MBio* **11**, e02415–e02419 (2020).
28. van Opijnen, T., Bodi, K. L. & Camilli, A. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat. Methods* **6**, 767–772 (2009).
29. Hutchison, C. A. et al. Polar effects of transposon insertion into a minimal bacterial genome. *J. Bacteriol.* **201**, e00185–19 (2019).
30. van Opijnen, T., Lazinski, D. W. & Camilli, A. Genome-wide fitness and genetic interactions determined by Tn-seq, a high-throughput massively parallel sequencing method for microorganisms. *Curr. Protoc. Mol. Biol.* **106**, 7.16.1–7.16.24 (2014).
31. van Opijnen, T. & Camilli, A. Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. *Nat. Rev. Microbiol.* **11**, 435–442 (2013).
32. Yasir, M. et al. TraDIS-Xpress: a high-resolution whole-genome assay identifies novel mechanisms of triclosan action and resistance. *Genome Res* **30**, 239–249 (2020).
33. Vigouroux, A., Oldewurtel, E., Cui, L., Bikard, D. & Teeffelen, S. Tuning dCas9's ability to block transcription enables robust, noiseless knockdown of bacterial genes. *Mol. Syst. Biol.* **14**, e7899 (2018).
34. Cui, L. et al. A CRISPRi screen in *E. coli* reveals sequence-specific toxicity of dCas9. *Nat. Commun.* **9**, 1912 (2018).
35. Lawson, M. J. et al. In situ genotyping of a pooled strain library after characterizing complex phenotypes. *Mol. Syst. Biol.* **13**, 947 (2017).
36. Camsund, D. et al. Time-resolved imaging-based CRISPRi screening. *Nat. Methods* **17**, 86–92 (2020).
37. Shiver, A. L., Culver, R., Deutschbauer, A. M. & Huang, K. C. Rapid ordering of barcoded transposon insertion libraries of anaerobic bacteria. *Nat. Protoc.* **16**, 3049–3071 (2021).
38. Baym, M., Shaket, L., Anzai, I. A., Adesina, O. & Barstow, B. Rapid construction of a whole-genome transposon insertion collection for *Shewanella oneidensis* by Knockout Sudoku. *Nat. Commun.* **7**, 13270 (2016).
39. Zhu, L. J., Holmes, B. R., Aronin, N. & Brodsky, M. H. CRISPRseek: a Bioconductor package to identify target-specific guide RNAs for CRISPR-Cas9 genome-editing systems. *PLoS One* **9**, e108424 (2014).
40. Drost, H.-G. & Paszkowski, J. Biomart: genomic data retrieval with R. *Bioinformatics* **33**, 1216–1217 (2017).
41. Blin, K., Pedersen, L. E., Weber, T. & Lee, S. Y. CRISPy-web: an online resource to design sgRNAs for CRISPR applications. *Synth. Syst. Biotechnol.* **1**, 118–121 (2016).
42. Doench, J. G. et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).
43. Sanson, K. R. et al. Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nat. Commun.* **9**, 5416 (2018).
44. van Gestel, J., Hawkins, J. S., Todor, H. & Gross, C. A. Computational pipeline for designing guide RNAs for mismatch-CRISPRi. *STAR Protoc.* **2**, 100521 (2021).
45. Spoto, M., Guan, C., Fleming, E. & Oh, J. A universal, genomewide GuideFinder for CRISPR/Cas9 targeting in microbial genomes. *mSphere* **5**, e00086–20 (2020).
46. Keller, L. E., Rueff, A.-S., Kurushima, J. & Veening, J.-W. Three new integration vectors and fluorescent proteins for use in the opportunistic human pathogen *Streptococcus pneumoniae*. *Genes* **10**, 394 (2019).
47. Sorg, R. A., Kuipers, O. P. & Veening, J.-W. Gene expression platform for synthetic biology in the human pathogen *Streptococcus pneumoniae*. *ACS Synth. Biol.* **4**, 228–239 (2015).

48. Read, A., Gao, S., Batchelor, E. & Luo, J. Flexible CRISPR library construction using parallel oligonucleotide retrieval. *Nucleic Acids Res.* **45**, e101 (2017).
49. Joung, J. et al. Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nat. Protoc.* **12**, 828–863 (2017).
50. Bokulich, N. A. et al. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat. Methods* **10**, 57–59 (2013).
51. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
52. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
53. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
54. Li, W. et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* **15**, 554 (2014).
55. Winter, J. et al. CRISPRAnalyzeR: interactive analysis, annotation and documentation of pooled CRISPR screens. Preprint at *bioRxiv* <https://doi.org/10.1101/109967> (2017).
56. Whatmore, A. M., Barcus, V. A. & Dowson, C. G. Genetic diversity of the streptococcal competence (com) gene locus. *J. Bacteriol.* **181**, 3144–3154 (1999).
57. Zhu, A., Ibrahim, J. G. & Love, M. I. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics* **35**, 2084–2092 (2019).
58. Pozzi, G. et al. Competence for genetic transformation in encapsulated strains of *Streptococcus pneumoniae*: two allelic variants of the peptide pheromone. *J. Bacteriol.* **178**, 6087–6090 (1996).
59. Kurushima, J. et al. Unbiased homeologous recombination during pneumococcal transformation allows for multiple chromosomal integration events. *eLife* **9**, e58771 (2020).
60. Sorg, R. A. et al. Collective resistance in microbial communities by intracellular antibiotic deactivation. *PLOS Biol.* **14**, e2000631 (2016).

Acknowledgements

Work in the Veening lab is supported by the Swiss National Science Foundation (SNSF) (project grants 31003A_172861 and 310030_192517), SNSF JPIAMR grant (40AR40_185533), SNSF NCCR 'AntiResist' (51NF40_180541) and ERC consolidator grant 771534-PneumoCaTChER.

Author contributions

X.L. developed the experimental procedures. V.d.B. developed the computational/analysis procedures. A.M.B. developed 2FAST2Q and the Python code for the computational pipelines. J.W.V. conceptualized this study. V.d.B. and X.L. wrote the original manuscript draft with input from A.M.B. V.d.B., X.L., A.M.B. and J.W.V. revised and edited to obtain the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41596-021-00639-6>.

Correspondence and requests for materials should be addressed to Jan-Willem Veening.

Peer review information *Nature Protocols* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 5 February 2021; Accepted: 24 September 2021;

Published online: 07 January 2022

Related links

Key references using this protocol

Liu, X. et al. *Cell Host Microbe* **29**, 107–120.e6 (2021): <https://doi.org/10.1016/j.chom.2020.10.001>

Liu, X. et al. *Mol. Syst. Biol.* **13**, 931 (2017): <https://doi.org/10.15252/msb.20167449>

Key data used in this protocol

Liu, X. et al. *Cell Host Microbe* **29**, 107–120.e6 (2021): <https://doi.org/10.1016/j.chom.2020.10.001>