nature protocols

PROTOCOL

https://doi.org/10.1038/s41596-021-00582-6

Check for updates

SorTn-seq: a high-throughput functional genomics approach to discovering regulators of bacterial gene expression

Leah M. Smith[1], Simon A. Jackson[1,2], Paul P. Gardner[2,3,4] and Peter C. Fineran[1,2,4]✉

We recently developed a high-throughput functional genomics approach, named 'SorTn-seq', to identify factors affecting expression of any gene of interest in bacteria. Our approach facilitates high-throughput screening of complex mutant pools, a task previously hindered by a lack of suitable techniques. SorTn-seq combines high-density, Tn*5*-like transposon mutagenesis with fluorescence-activated cell sorting of a strain harboring a promoter-fluorescent reporter fusion, to isolate mutants with altered gene expression. The transposon mutant pool is sorted into different bins on the basis of fluorescence, and mutants are deep-sequenced to identify transposon insertions. DNA is prepared for sequencing by using commercial kits augmented with custom primers, enhancing ease of use and reproducibility. Putative regulators are identified by comparing the number of insertions per genomic feature in the different sort bins, by using existing bioinformatic pipelines and software packages. SorTn-seq can be completed in 1–2 weeks and requires general microbiology skills and basic flow cytometry experience.

Introduction

Investigating bacterial phenotypes and regulatory networks requires an understanding of underlying genetic components and their contributions. Targeted mutagenesis is a useful technique for interrogating bacterial networks but often requires prior knowledge of the system and developed genetic tools for the organism in question. In contrast, transposon mutagenesis is a broadly applicable, semi-random technique for mutating bacterial chromosomes[1]. Transposons are self-mobilizable genetic elements that often confer loss-of-function mutations when inserted into a gene or regulatory region. Type II DNA transposons disseminate through 'cut and paste' mechanisms, during which the transposon self-excises from host DNA and integrates at a new location[2]. Integration of some transposons, such as the bacterially derived Tn*5*, is largely random, whereas others, like the eukaryotic-derived *Himar1*/*Mariner*, require certain sequence-recognition motifs[3–6].

Transposon mutagenesis has been coupled with deep sequencing to study the essentiality and fitness contributions of genes within bacteria under different conditions. There are several variations of transposon insertion sequencing (TIS) that differ mainly in the transposon type used and the enrichment of transposon-chromosomal junctions during sequencing library preparation. Among the most widely used methods are the *mariner* transposon-based approaches, transposon sequencing (Tn-seq)[7] and insertion sequencing (INSeq)[8], as well as the transposon directed insertion sequencing (TraDIS)[9,10] technique, which offers flexibility in transposon type. Furthermore, several unnamed variants of TIS that use different transposons and enrichment schemes have been developed[6].

One challenge of using TIS-based approaches for questions beyond gene essentiality and fitness is the development of suitable mutant library screening procedures, because most traditional TIS experiments rely on negative selection screens. Recently, TIS has been coupled with various screening methods to study biological questions ranging from gene function and regulation to host-pathogen interactions[6,11,12]. However, many screening procedures have been designed to answer very specific biological questions, thereby precluding their broad application to different genetic systems or organisms. Hence, we developed SorTn-seq (pronounced 'sort 'n seek'), which couples a versatile transposon library screening approach with deep sequencing, to identify regulators of target gene

[1]Department of Microbiology and Immunology, University of Otago, Dunedin, New Zealand. [2]Genetics Otago, University of Otago, Dunedin, New Zealand. [3]Department of Biochemistry, University of Otago, Dunedin, New Zealand. [4]Bio-Protection Research Centre, University of Otago, Dunedin, New Zealand. ✉e-mail: peter.fineran@otago.ac.nz

expression. SorTn-seq was developed and applied to systematically uncover regulators of type III-A CRISPR–Cas (*csm*) expression in *Serratia* sp. ATCC 39006[13] (hereafter called *Serratia*). While we were interested in defining the regulatory networks controlling bacterial adaptive immunity, this method can be easily customized to identify factors modulating the expression of any gene of interest in organisms with suitable transposon mutagenesis and fluorescent reporter availability.

SorTn-seq combines fluorescent reporters, transposon mutagenesis, FACS and TIS to identify regulators of target gene expression (Fig. 1). Flow cytometry and FACS facilitate the isolation of single cells on the basis of size, internal complexity or fluorescence parameters[14]. In SorTn-seq, FACS facilitates the rapid enrichment of bacterial cells with altered fluorescent reporter activity from within a saturated transposon mutant library. Sample libraries for deep sequencing of the transposon insertion sites are then prepared by using a protocol that is based mainly on modifications to commercially available kits. We also provide a bioinformatic data analysis protocol, which uses several established tools[10,15] to map the locations of transposon insertions and predict regulators of the gene of interest.

An overview of the SorTn-seq method is outlined in Fig. 1. First, a promoter-fluorescent reporter fusion is constructed to measure the activity of any gene of interest. This fusion may report on gene transcription, translation or transcription and translation (Fig. 1a), depending on the genetic elements incorporated[16]. Cells harboring this reporter are then subjected to random Tn*5*-based transposon mutagenesis to generate a mutant library (Fig. 1b). To identify mutations influencing target gene activity (i.e., insertions in activators and/or repressors), an aliquot of the transposon mutant library is grown and ~2.0×10^7 cells are sorted (binned) on the basis of fluorescence intensity. Cells with altered gene activity are thereby enriched within the low or high expression bins relative to the bin that is depleted of potential regulators (i.e., cells with intermediate expression) (Fig. 1c). FACS enrichment of mutant library aliquots is performed in triplicate. After each experiment, DNA is extracted from the different bins, as well as from an unsorted control sample (input), which provides information on gene essentiality and library diversity but is not required (Fig. 1d). TIS libraries are then prepared and sequenced (Fig. 1e), and the resulting reads are mapped to the chromosome by using the TraDIS[10] pipeline to locate transposon insertion sites within each bin (Fig. 1f). Custom R scripts are used to map insertion sites to genomic features and generate counts tables (reads and insertions) per feature. Potential regulators are identified using the edgeR[15] package for differential expression analysis, through comparison of the number of unique insertions within a gene or feature between the low or high and the depleted activity bins.

Comparison with existing methods

The countless, unexplored aspects of bacterial physiology have led to the convergence of several new techniques that use various combinations of reporter systems, transposon mutagenesis, TIS and FACS. The usefulness of SorTn-seq is discussed below, with respect to methods using related techniques.

Early FACS-based screens of bacterial mutant libraries required recovery and sequencing of individual clones. For example, a combination of transposon mutagenesis and FACS was used to find regulators of a virulence factor promoter-GFP fusion in *Vibrio cholerae*[17]. Individual cells were recovered, and transposon insertions were determined by using Sanger sequencing. Similarly, a FACS-based screening approach was used alongside chemical mutagenesis to isolate *Corynebacterium glutamicum* mutants with altered levels of L-lysine production[18]. Cells were sorted on the basis of fluorescence generated by a lysine biosensor, after which individual clones were Sanger-sequenced to identify causative mutations. Although the FACS-based screening methods used in both studies are high throughput, the requirements of individual clone isolation and sequencing is labor intensive and restricts broad utility.

Spinoffs of the TraDIS[9,10] technique have been used to study drug resistance and virulence determinants in gram-negative pathogens[19,20]. Genes comprising and regulating drug efflux systems in *Acinetobacter baumannii* were identified by using TraDISort, an approach combining transposon mutagenesis, FACS and TIS[19]. Mutants deficient in efflux activity were sorted on the basis of levels of intracellular ethidium bromide accumulation and then sequenced to identify insertion sites. Another approach (Density-TraDISort) was used to identify regulators of capsule formation in *Klebsiella pneumoniae* by using density-gradient centrifugation to sort transposon mutant libraries on the basis of capsule biosynthesis[20]. Although TraDISort and density-TraDISort are useful techniques for elucidating the genetic components of their respective pathways, they were designed to explore very

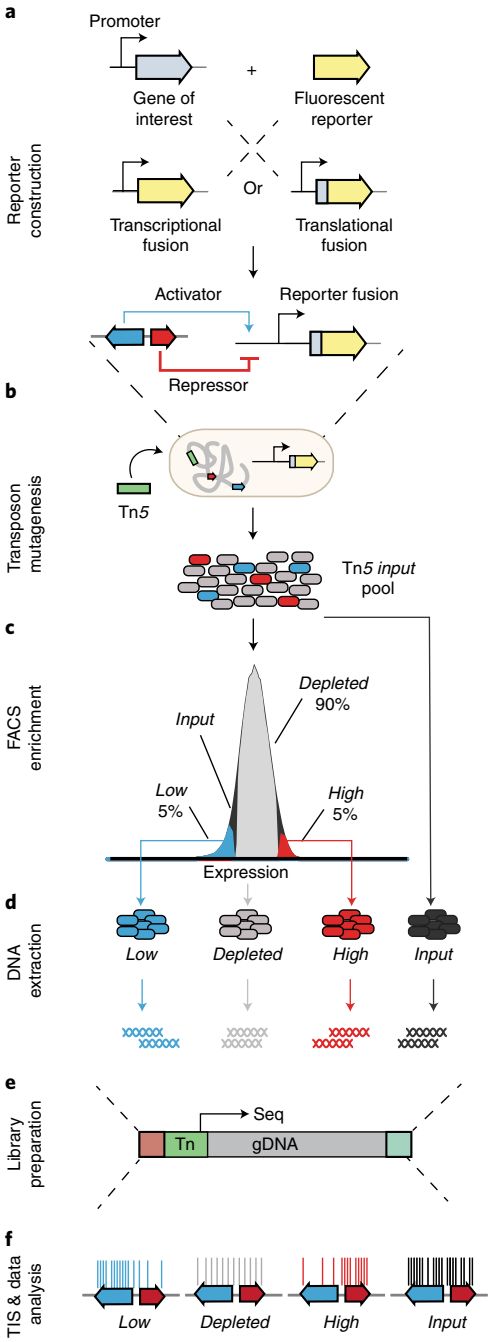


Fig. 1 | SorTn-seq protocol overview. a, A promoter-fluorescent reporter (protein or RNA) fusion reports on the transcription and/or translation of a gene of interest. **b** and **c**, Cells harboring a reporter fusion are mutagenized by using Tn*5* (**b**; Steps 1–33) and then screened for activity through FACS (**c**; Steps 34–40). Mutants are sorted on the basis of gene activity levels (low, high or depleted), after which DNA is extracted from each sort bin and an unsorted control (input) (**d**; Steps 41–50). **e**, DNA is then used to prepare TIS libraries (Steps 51–105), which are sequenced from the transposon into the genomic DNA (gDNA) junction (Steps 125–129). **f**, After sequencing, reads are mapped to the chromosome to identify the location of transposon insertions (vertical bars) within each bin (Steps 130–137). The number of unique insertions within a gene or feature is compared between the low or high and depleted bins to identify potential regulators (Steps 138–141). Figure adapted from ref. [13].

specific phenotypes. Likewise, these approaches do not use reporter gene technology, which limits their wider applicability to other research questions.

Mutagenesis and FACS-based approaches have also been used to investigate bacterial predator-prey dynamics[21], as well as to identify factors that influence antibiotic susceptibility and virulence in

different human pathogens[21–23]. Tn-FACSeq was developed to identify mutants of the bacterial predator *Bdellovibrio bacteriovorus* that are deficient in forming attachment to *V. cholerae* prey[21]. Genes important for attachment were identified through Tn-seq, after using FACS to isolate *V. cholerae* cells, which either harbored or were deficient in fluorescently labeled *B. bacteriovorus*. In *Mycobacterium tuberculosis*, a *mariner* transposon mutant library was stained and sorted by using calcein, a fluorescent molecule that acts as a predictor of antibiotic susceptibility. Deep sequencing of sorted cells allowed researchers to identify a link between phenotypic heterogeneity and antibiotic susceptibility[22]. Although informative, the use of calcein as a reporter limits the applications of this technique to studies of a similar nature. Another approach, fluorescence-activated sorting of transposon mutants coupled with insertion site sequencing, was used to isolate mutations in *Salmonella enterica* that influence the expression of typhoid toxin within infected epithelial cells[23].

A recently developed technique, droplet-Tn-seq allows for the isolation of individual *mariner* transposon mutants within droplets on microfluidic chips before sequencing[24]. Segregated mutant growth within droplets and comparison of mutant frequencies before and after growth facilitates the assessment of single-cell fitness, independently of population-level effects. Although this method was developed to study the fitness of mutants grown in isolation, the authors also demonstrated that GFP could be detected by flow cytometry, creating the possibility for sorting droplets with altered fluorescence. One consideration of the droplet-Tn-seq approach is accessibility, because the cost and expertise required for microfluidic chip design and fabrication may not be feasible for many research groups.

In experiments using Tn-seq[7], INSeq[8,25] or variants thereof, mutagenesis with a modified *mariner* transposon limits transposition to AT sites within the chromosome[3,4]. In screens based on the INSeq approach, a modified *mariner* transposon containing an *MmeI* restriction site facilitates sequencing library construction[25]. After *MmeI* DNA digestion, 20 nt of flanking chromosomal DNA is incorporated into constructs for sequencing. Although 20-nt reads are sufficient for mapping in most instances, the short reads generated from these experiments can pose a challenge for identifying insertions within orthologs or repetitive regions within a genome. However, library preparation with other *mariner* transposons does not require *MmeI* digestion; therefore, the length of flanking DNA can be longer in those instances. Tn*5* transposons, like that used during SorTn-seq, show less bias toward specific integration motifs than *mariner*, thereby facilitating very high insertion densities[5,6,9,26,27]. High-density transposon saturation in SorTn-seq allows comparison of the unique number of insertions within genes and other genomic features between different treatments (sorted cell bins) to identify regulators. Basing analysis on unique insertions rather than read counts is probably more robust, because the latter can be influenced by amplification bias during the sequencing library preparation stage[28]. In addition, during SorTn-seq sequencing library preparation, random DNA fragmentation allows larger regions of chromosomal DNA to be incorporated into final library products, which may help alleviate potential problems of correctly identifying the origin of shorter reads.

The SorTn-seq sequencing library preparation method uses commercial kits for sequencing library construction, abrogating the need for custom adaptor and index primer design, such as those used during TraDIS[9,10]-based approaches. In addition, our nested PCR scheme using biotinylated product capture ensures maximal transposon-chromosome junction enrichment. Illumina P5/P7 sequences, which are required for library–flow cell hybridization, are incorporated after transposon enrichment. This minimizes the generation of nonspecific products that compete for clustering space on the flow cell. Using this approach in *Serratia*, an average of 94% of all reads (averaged across different sequencing libraries) harbored the specified transposon 'tag'[13], indicating their origin from bona fide transposon-chromosomal junctions. However, the presence of the transposon tag results in limited nucleotide diversity, which can pose a challenge for the Illumina sequencing platform[29]. To circumvent these issues, libraries are pooled with the commercially available PhiX control library and loaded at decreased concentrations to ensure optimal flow cell cluster generation for lower-diversity samples. Although our approach generates less raw sequencing data than a high-diversity run, it can be performed without the required 'dark cycles' of the TraDIS[10] technique, which is currently unsupported by Illumina and many sequencing facilities.

Experimental design

SorTn-seq requires that both fluorescent reporters and transposon mutagenesis protocols be established in the organism of interest. Transposon mutagenesis has been successfully used in many

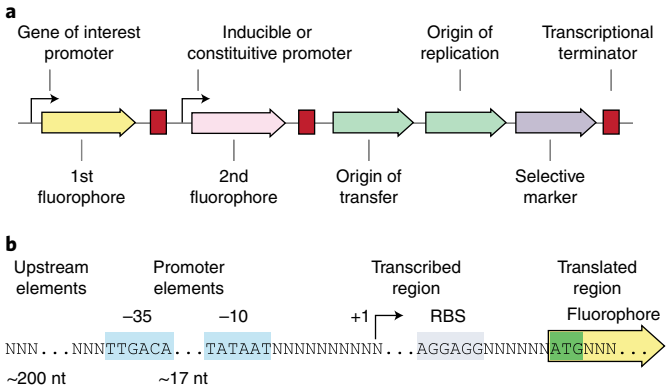


Fig. 2 | Design of the fluorescent reporter plasmid used during SorTn-seq. a, The promoter region for the gene of interest drives expression of the primary reporter (1st fluorophore), which could be a fluorescent protein or RNA. A second fluorescent reporter (2nd fluorophore), driven by an inducible or constitutive promoter, aids in flow cytometry gating and serves as a proxy for plasmid copy number (Extended Data Fig. 1). An origin of transfer facilitates mobilization of the plasmid via conjugation, and the origin of replication enables replication within the donor and recipient. A selective marker (e.g., an antibiotic resistance gene) is required for transconjugant selection and reporter maintenance. The inclusion of transcriptional terminators (red) minimizes transcriptional read-through between plasmid components. **b**, The promoter-fluorophore fusion should contain RNA-polymerase recognition sequences from the gene of interest (−35 and −10, blue) through to potential transcription initiation sites (+1, arrow). A sequence upstream of the promoter should also be included, because it may contain binding sites for regulatory elements. To identify factors that influence translation, the native ribosome binding site (RBS, gray) start codon (ATG, green) and first few amino acids (optional) of the gene of interest should be fused to the fluorescent protein.

bacterial species (for a comprehensive summary, see ref. [6]), and fluorescent reporters (proteins and RNAs) are constantly being evolved with useful characteristics and improved host ranges[30–32]. After transposon mutant library construction, standard FACS is performed to enrich mutants of interest. Finally, SorTn-seq requires custom primers specific to the transposon used for generating TIS libraries. The following sections address key elements required for the successful design of fluorescent reporters, transposon mutagenesis experiments and custom primers, as well as tips for adapting this protocol to the organism of interest.

Fluorescent reporter design

By coupling a fluorescent reporter with flow cytometry, the expression of any gene of interest can be measured in individual cells. SorTn-seq uses a plasmid-based fluorescent reporter, which provides the advantage of an increased signal-to-noise ratio for low- to medium-expressed genes. The reporter may be adapted for chromosomal integration if the gene is highly expressed. Another advantage of plasmid-based reporters is that they are more amenable in less genetically tractable organisms, where tools for chromosomal manipulation are not well developed.

The general design of the reporter plasmid is outlined in Fig. 2a. The promoter of the gene of interest is fused to the primary fluorescent reporter (1st fluorophore; protein or RNA), and cells are sorted on the basis of the primary reporter fluorescence levels. A secondary fluorescent reporter (2nd fluorophore) is under the control of an inducible or constitutive promoter. This additional fluorophore enhances the gating strategy applied to cells during the sorting process to reduce noise (Extended Data Fig. 1). In addition, this secondary reporter can be used as a proxy to assess mutation-induced fitness effects or alterations in reporter plasmid copy number. Fluorophores should be chosen on the basis of compatibility with the available flow cytometry equipment as well as detectability and intensity within the organism, with considerations placed on fluorophore 'half-life' (turnover and stability) as well as potential toxicity[32,33].

Ideally, the two fluorophores chosen will have minimal spectral overlap (i.e., their emission and excitation wavelengths are discrete); however, mathematics-based 'compensation' tools can be applied to alleviate this requirement[34]. Where compensation is used, two control versions of the reporter plasmid should be constructed, each containing only one fluorophore. In addition, some bacteria exhibit intrinsic fluorescence, which should be assessed during the reporter design stage. For example, *Serratia* produces the pigmented secondary metabolite prodigiosin, which has fluorescent

properties[35,36]. Therefore, SorTn-seq in *Serratia* was performed in a mutant deficient in prodigiosin production to minimize interference with the detection of fluorescent reporter proteins[13].

For bacteria that are not readily transformed, an origin of transfer gene (*oriT*) can be used to transfer the reporter plasmid via conjugation. The construct must also possess an origin of replication (*oriV*), which should be selected based on (i) compatibility with the donor and recipient, (ii) the desired copy number of the plasmid and (iii) compatibility with downstream applications (e.g., using the reporter in conjunction with other plasmids). Depending on the strength of transcription from the gene-of-interest promoter, an origin of replication facilitating low to medium plasmid copy number is probably sufficient (e.g., RK2, p15a and ColE1 origins[37]) and will require experimental validation (Fig. 2a).

The most crucial aspect of the plasmid design is the fluorescent protein reporter fusion. At a minimum, this fusion should contain promoter elements (−35 and −10), a transcriptional start site (+1), ribosome binding site (RBS) and start codon (ATG) (Fig. 2b). The fluorescent reporter plasmid used during *csm* SorTn-seq in *Serratia*[13] is outlined in Extended Data Fig. 2. For a fusion that reports on gene transcription and translation, the native start codon and upstream RBS region should be fused to a fluorescent protein (so that the native ATG comprises the first amino acid of the fluorophore). In addition, a small number of amino acids downstream of the start codon may be included, because these regions can be important for small RNA-mediated post-transcriptional regulation[38]. For a fusion reporting only on transcriptional activity, the 5′ untranslated region (after the transcriptional start site) can be replaced with an artificial RBS/start codon region to optimize translational efficiency. Alternatively, a fluorescent RNA can serve as the reporter fused to promoter elements, which does not require an RBS/start codon[39,40]. In bacteria, some transcriptional regulators act upstream of promoter elements[41]. For this reason, it is recommended to include ≥200 nt upstream of the promoter. However, the length of the sequence included may require optimization based on the gene/organism in question and knowledge of the promoter location.

Transposon mutagenesis design

The generation of a transposon mutant library will require optimization based on the organism and can vary in (i) the type of transposon used and (ii) the method of transposon delivery. The SorTn-seq protocol has been designed and tested by using a Tn*5*-derived transposon introduced from a suicide delivery vector (Extended Data Fig. 3a). Transposon delivery to the recipient occurs via conjugation with an *Escherichia coli* auxotrophic donor (ST18[42,43]), thereby simplifying the transconjugant selection process. For example, in *Serratia*, we find conjugation to be more efficient than transformation. After conjugation, transposon mutants are enriched through liquid selection, thus minimizing the time and resources required to plate and recover individual colonies. The transposon we used (Tn-DS1028*uidA*Km) encodes kanamycin resistance and an R6Kγ origin of replication (Extended Data Fig. 3b) capable of initiating replication only in *pir*+ strains. Following mobilization of the delivery vector into a *pir*− recipient, the plasmid-encoded transposase facilitates the excision of the transposon and integration into the host. The delivery vector is then lost because of its inability to replicate in the recipient.

During transposon library construction, we performed mutagenesis on *Serratia* already harboring the reporter plasmid. This approach was selected to reduce potential library diversity bottlenecks, which could occur when introducing a reporter plasmid into a mutant pool. By using our approach, the transposon can insert within the fluorophore, and those cells may subsequently be enriched during sorting. However, we analyzed transposon insertions only within the *Serratia* chromosome to identify potential regulators (Supplementary Table 1). For our dataset, only 0.73% of reads mapped to the reporter plasmid (as averaged across all low, high, depleted and input replicates; Supplementary Table 2), leaving the vast majority of reads to identify chromosomal mutations of interest.

Use of transposon mutagenesis and fluorescent reporters in other organisms

Although we chose a Tn*5*-transposon derivative to achieve high insertion numbers, mutagenesis can also be performed by using other transposons and delivery mechanisms (such as transformation or phage delivery systems), which may offer greater compatibility with the organism of interest. A recent review by van Opijnen and Levin[6] includes a helpful list of both the organism and transposon type used in nearly 300 transposon mutagenesis studies. The exact Tn*5*-transposon delivery system used during *Serratia* SorTn-seq[13] has been successfully used in both *Pectobacterium atrosepticum*[44] and *Pseudomonas syringae*[45], while variants of the transposon carrying different markers have been used

to mutagenize *Chromobacterium, Citrobacter, Dickeya, Erwinia* and other *Pectobacterium* and *Serratia* species[46].

Fluorescent proteins can be introduced and maintained in many different bacteria, including gram-negative bacteria[30,47], rhizobia[48], lactic acid bacteria[49] and other gram-positive species[50,51]. The eYFP and mCherry fluorescent reporters used during SorTn-seq are part of the Rainbow Vector series for broad-range host labeling[52], which have been validated in several gram-negative bacteria including *E. coli, Pseudomonas aeruginosa, K. pneumoniae*, *Burkholderia cepacia, Bordetella bronchiseptica* and *Shigella flexneri*. In addition, our group has successfully used a similar eYFP and mCherry reporter plasmid in *P. atrosepticum* to study anti-CRISPR promoter regulation[53].

TIS: overview and primer design

After reporter design, mutagenesis and FACS, the DNA from sorted cells is isolated to prepare TIS libraries for Illumina sequencing. SorTn-seq uses a combination of commercially available kits with custom enrichment and sequencing primers. Our sequencing library preparation protocol enriches transposon-chromosomal junctions by using a nested PCR approach (Fig. 3a). After DNA fragmentation/end repair and adaptor ligation, a biotinylated primer is used to generate transposon-specific PCR products that are subsequently captured with streptavidin beads. A second PCR enrichment step adds sequences required for Illumina flow cell binding (P5/P7 sequences) as well as a unique index to each sample. After final purification, library quality is assessed in several ways. The library concentration is measured by using Qubit fluorometric quantitation, DNA size distribution is assessed by using a Bioanalyzer and the presence of transposon junctions and sequences required for Illumina flow cell binding is determined through qPCR (Fig. 3a). Samples are then pooled, and PhiX control library is added to increase nucleotide diversity. Sequencing is performed on an Illumina MiSeq (or equivalent) by using a custom Read 1 primer that hybridizes to the end of the transposon, generating reads that begin with a 12-nt transposon tag (Fig. 3b). This tag facilitates the filtering of sequencing reads derived from bona fide transposon-chromosomal junctions.

A schematic of the primer setup used during construction of Tn-DS1028*uidA*Km sequencing libraries is shown in Fig. 3c. The enrichment and sequencing primers are tailored to the transposon used and should include the key elements outlined in Table 1. Enrichment primers were designed so that the melting temperature (T_m) of the region hybridizing with the transposon was ≥64 °C (as assessed by using OligoEvaluator by Sigma-Aldrich). In addition, the transposon sequencing primer should have a T_m similar to that of the Illumina Read 1 primer (77 °C). Primer designs should also be assessed for potential secondary structure and T_m compatibility.

Data analysis

After sequencing, potential regulators of the gene of interest are identified by using tools from the Bio-TraDIS[10] pipeline and edgeR[15] software (Fig. 4). Briefly, the FASTQ files generated after sequencing are assessed for quality/adaptor contamination. The Bio-TraDIS pipeline is then used to filter reads containing the transposon tag and to map these transposon-derived reads to the chromosome. This generates a summary file with mapping and insertion statistics, as well as sequence alignment files. In R, alignment files are used to generate tables of reads and insertions per nucleotide for all samples. The insertions at each nucleotide position are assigned to genomic features, as specified by a table listing the coordinates of genomic features in the organism of interest. The number of unique insertions per feature (for each sample) serves as the input for analysis by edgeR[15], which fits the data to a negative binomial distribution and performs a differential expression analysis. Here, replicates of the low and high expression libraries are compared against the depleted libraries, which serve as the control. The edgeR output contains the $\log_2$ fold change (or 'enrichment') of each feature in the low/high as compared to the depleted libraries and *P* values (adjusted for multiple hypothesis testing) to assess statistical significance (Fig. 4). Putative regulators are those with a positive $\log_2$ fold change and significant (adjusted) *P* value, indicating that mutations in these features were enriched in the altered fluorescence bins.

Limitations

The main tenets of SorTn-seq require that an organism be able to: (i) harbor fluorescent reporters and (ii) be mutated by using transposons. If the gene of interest is expressed at very low levels, there may be insufficient fluorescence to distinguish changes in expression/intensity from background noise. In these cases, it may be possible to 'boost' the fluorescent signal by using tandem repeats of the

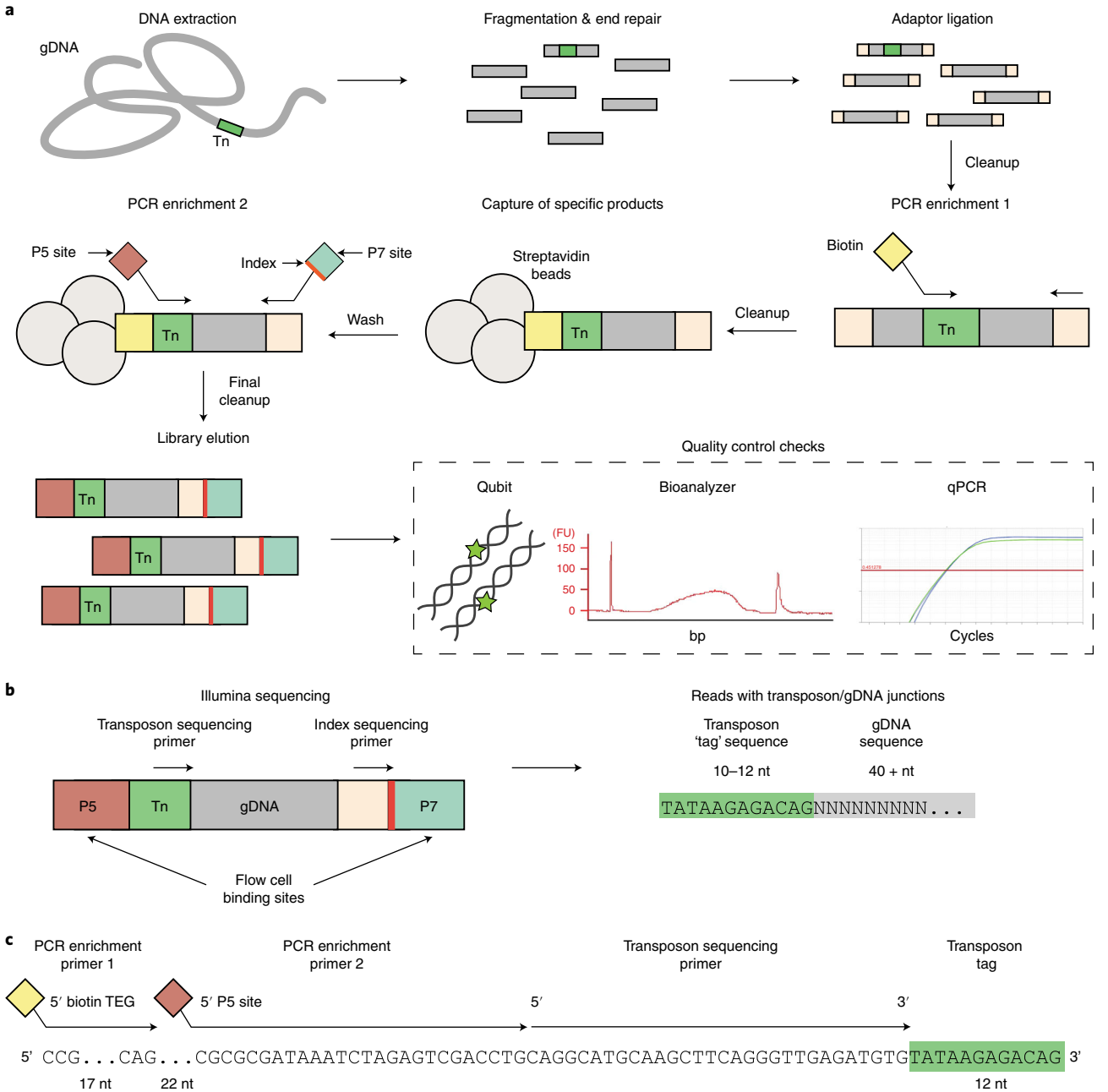


Fig. 3 | SorTn-seq TIS workflow. a, DNA is isolated from sorted cells, after which fragmentation/end repair and adaptor ligation are performed. A biotinylated primer is used to enrich transposon (Tn) sequences during PCR enrichment 1. Biotinylated products are captured with streptavidin beads, and a second PCR enrichment adds Illumina flow cell hybridization sequences (P5/P7 sites) as well as a unique index to each sample. After final purification, the libraries are assessed for quality. **b**, Final libraries are pooled and sequenced by using a custom Read 1 primer that binds to the end of the transposon, generating a transposon-derived tag sequence on each read. The standard Illumina primer mix is used for index sequencing. **c**, Position of library preparation and sequencing primers along the Tn-DS1028*uid*AKm transposon. PCR enrichment primer 1 contains a 5′ biotin molecule and a 15-carbon tetraethylene glycol (TEG) spacer arm to minimize steric hindrance during PCR enrichment. In the second round of amplification, PCR enrichment primer 2 binds 22 nt downstream of the primer 1 binding site, further preventing steric hindrance as templates are bound by streptavidin beads during PCR. The transposon sequencing primer hybridizes downstream of the PCR enrichment primer 2 binding site and immediately upstream of the end of the transposon. Thus, each read deriving from a transposon-chromosomal junction begins with a 12-nt transposon tag sequence. The transposon sequencing primer binding site may also overlap with PCR enrichment primer 2.

fluorescent protein[54], but this would require optimization. Alternatively, a higher-copy-number plasmid (e.g., pUC) could be used. In addition, mutations that can be complemented in trans (i.e., mutations that can be rescued by products provided by other members of the population) may be difficult to identify.

Table 1 | Primers and elements required for preparation of SorTn-seq transposon sequencing libraries

Name	Binding site	Required elements/description	Notes
PCR enrichment primer 1	Binds within the transposon (furthest upstream)	5′ biotin with tetraethylene glycol (TEG) spacer arm to minimize steric hindrance	Pair with adaptor-specific primer 1
5′ /BiotinTEG/TCATCTGCAGCCGGGAATTCTCATGTTTGACA*G 3′			
Adaptor-specific primer 1	Binds to the NEBNext adaptor for Illumina	Compatibility with the NEBNext adaptor for Illumina	Pair with PCR enrichment primer 1; NEBNext adaptor for Illumina included with NEBNext multiplex oligos for Illumina (New England Biolabs, cat. no. E7335)
5′ GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T 3′			
PCR enrichment primer 2	Binds within the transposon (downstream of PCR enrichment primer 1)	5′ Illumina P5 sequence; binds ~20 nt downstream of the PCR enrichment primer 1 binding site to avoid steric hindrance	Pair with index primer; P5 site allows hybridization to the Illumina flow cell
5′ AATGATACGGCGACCACCGAGATCTACACCGCGCGATAAATCTAGAGTCGACCT*G 3′			
Index primer (NEBNext multiplex oligos for Illumina)	Binds to the sequence introduced by adaptor-specific primer 1	5′ Illumina P7 sequence and index sequence. Use a unique indexing primer for each sample	Pair with PCR enrichment primer 2; P7 site allows hybridization to the Illumina flow cell; unique index allows sample multiplexing (New England Biolabs, cat. no. E7335)
Transposon sequencing primer	Binds within the transposon (downstream of PCR enrichment primer 2)	Orient binding so that the first 10–12 nt of sequencing reads originate from the transposon	This generates the transposon tag used to identify bona fide transposon insertion-derived reads
5′ CAGGCATGCAAGCTTCAGGGTTGAGATGTG 3′			
Illumina Read 1 primer	Binds the Illumina PhiX control library	The PhiX library is included during sequencing to increase diversity	Combine with the transposon sequencing primer to create a custom primer mix during Illumina sequencing
5′ ACACTCTTTCCCTACACGACGCTCTTCCGATCT 3′			

The sequences of primers used to generate Tn-DS1028*uid*AKm-based transposon sequencing libraries are shown, with relevant features underlined. All custom primers were ordered from Integrated DNA Technologies with HPLC purification. Index primers were ordered from New England BioLabs (cat. no. E7335), which includes the NEBNext adaptor for Illumina. * indicates a phosphorothioate bond to limit exonuclease digestion.

One caveat of all transposon mutagenesis screens is that mutants harboring insertions within essential genes are lost from the population. Therefore, regulators of the gene of interest that are essential for growth cannot be identified in this screen. However, transposon insertions in regulatory regions (e.g., promoters), which may alter gene expression levels rather than abrogate function, can still be identified by using SorTn-seq. Similarly, the TraDIS-Xpress method identified essential genes involved in triclosan resistance through analysis of intergenic insertions[55]. In TraDIS-Xpress, insertion of transposons (modified to contain an inducible promoter) upstream of open reading frames allowed researchers to control the expression levels of downstream genes[55]. Likewise, transposon library selection in liquid media and the outgrowth required for reporter transcription/translation may introduce biases in mutant pool composition through competition between constituent mutants with differing fitness. However, high-density transposon insertion and limited pool outgrowth may help to offset the loss of fitness-impaired mutants from within the population.

An important aspect of successfully using SorTn-seq is to generate a high-density transposon library. In *Serratia*, our transposon library contained 293,000 unique chromosomal insertions, or one insertion per 17 nt (on average)[13]. During transposon mutagenesis, we used colony counts to estimate the number of total (2.75×10^6) and unique (5.83×10^5) mutants in our library (Supplementary Tables 3–5). Some diversity is expected to be lost after library selection and outgrowth, but this pre-selection estimate can give an important indication of whether the library is likely to be sufficiently saturated. A higher number of unique transposon mutants facilitates prediction of a greater number of regulators (Supplementary Fig. 1). With lower transposon insertion density, only regulators whose disruption by mutagenesis results in the greatest expression differences are likely to be significantly enriched. In cases of lower insertion density, the data analysis protocol may also be run by using read counts (instead of solely unique insertion counts).

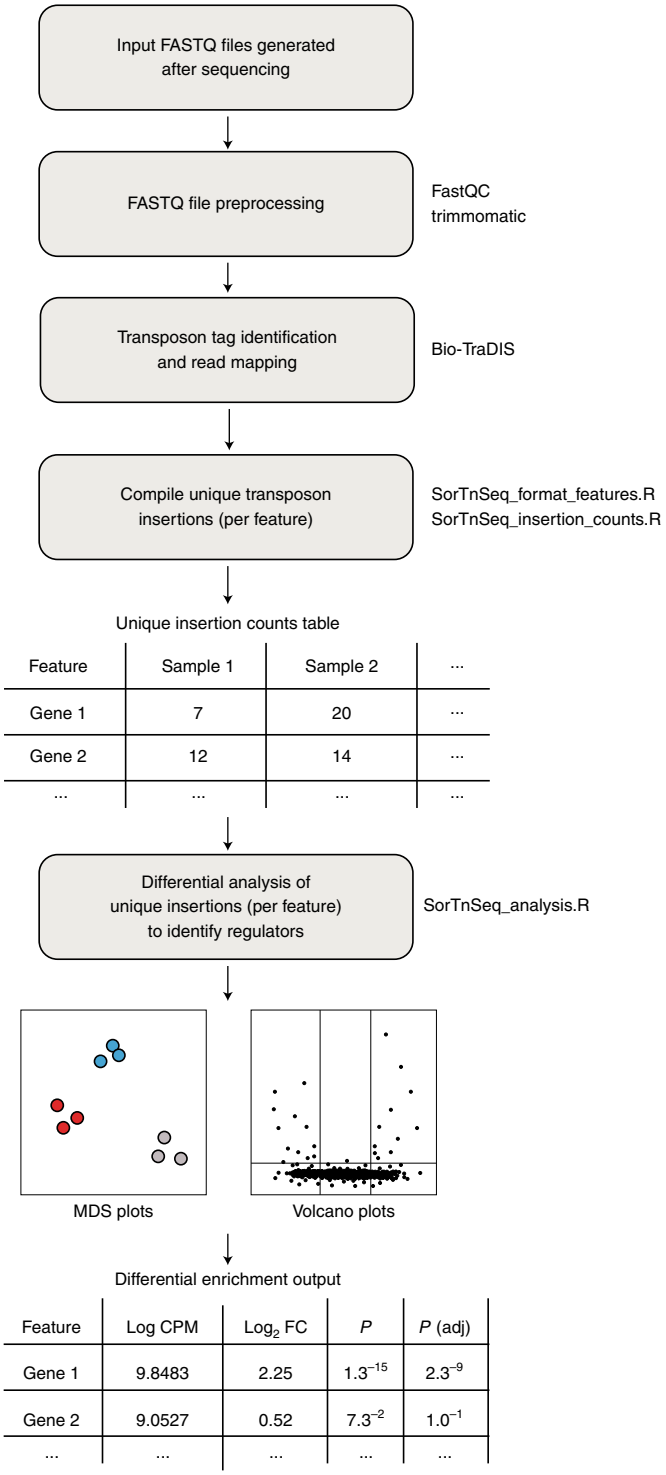


Fig. 4 | SorTn-seq data analysis workflow. Quality control of FASTQ files is performed, and reads are mapped to the reference genome by using the Bio-TraDIS[10] pipeline. In R, the number of unique insertions per feature is compiled and used to identify features enriched in the high/low bins by using edgeR[15]. For a detailed analysis workflow, including required file inputs and outputs, see Fig. 8. CPM, $\log_2$ counts per million; FC, $\log_2$ fold change; MDS, multidimensional scaling plot; *P*(adj), adjusted *P* value.

Finally, if the full SorTn-seq workflow cannot be used for a given system, parts of the technique may still be adaptable to different research questions. For example, our method of TIS library preparation is flexible and can be used for transposon sequencing experiments independently of

FACS. Likewise, our FACS procedure could be adapted to isolate cells from within other pools, such as the Keio collection[56], ASKA library[57] or CRISPRi libraries[58].

Expertise required

Most of the SorTn-seq procedure can be executed without specialist knowledge but does require the setup and optimization of FACS for bacterial cells and general familiarity with the preparation and quality control (QC) of Illumina sequencing libraries. Familiarity with basic UNIX commands and R, particularly the tidyverse package[59], is necessary for the data analyses.

Materials

▲ **CRITICAL** The strains, plasmids, primers and data files listed are those specific to the CRISPR-Cas SorTn-seq experiment described in Smith et al.[13]. Manufacturer names and catalog numbers are provided for nonstandard reagents and equipment. Strains and plasmids in this protocol are available upon request.

Biological materials

Transposon mutagenesis

- Recipient strain (*Serratia* PCF396; Supplementary Table 6) harboring fluorescent reporter (pPF1334; chloramphenicol resistant (Cm^R); Extended Data Fig. 2; Supplementary Table 7)
- Donor strain (*E. coli* ST18; 5-aminolevulinic acid hydrochloride (ALA) auxotroph, streptomycin resistant (Sm^R); Supplementary Table 6) harboring transposon delivery vector (pKRCPN2; backbone tetracycline resistant (Tc^R), transposon kanamycin resistant (Km^R); Extended Data Fig. 3; Supplementary Table 7)

FACS

- 'Unstained' control for spectral compensation (PCF396)
- 'Single stain' control (eYFP) for spectral compensation (PCF396 + pPF1307; Supplementary Table 7)
- 'Single stain' control (mCherry) for spectral compensation (PCF396 + pPF1438; Supplementary Table 7)

Reagents

Common reagents

- Kanamycin sulfate (Km) (Sigma-Aldrich, cat. no. 60615) **! CAUTION** Km has oral and dermal toxicity; use personal protective equipment including protective gloves, protective clothing and eye or face protection. Use in a well-ventilated area. Dispose of this material and its container in accordance with local regulations.
- Chloramphenicol (Cm) (Sigma-Aldrich, cat. no. C0378) **! CAUTION** Cm is a suspected carcinogen; use personal protective equipment including protective gloves, protective clothing and eye or face protection. Cm is flammable when dissolved in ethanol (EtOH). Use in a well-ventilated area and avoid ignition sources. Dispose of this material and its container in accordance with local regulations.
- Tetracycline (Tc) (Sigma-Aldrich, cat. no. 87128) **! CAUTION** Tc has oral, dermal and reproductive toxicity; use personal protective equipment including protective gloves, protective clothing and eye or face protection. Tc is flammable when dissolved in EtOH. Use in a well-ventilated area and avoid ignition sources. Dispose of this material and its container in accordance with local regulations.
- Streptomycin sulfate salt (Sm) (Sigma-Aldrich, cat. no. S6501) **! CAUTION** Sm has oral toxicity; use personal protective equipment including protective gloves, protective clothing and eye or face protection. Use in a well-ventilated area. Dispose of this material and its container in accordance with local regulations.
- ALA (Sigma-Aldrich, cat. no. A3785) **! CAUTION** ALA is combustible; use personal protective equipment including protective gloves, protective clothing and eye or face protection. Use in a well-ventilated area and avoid ignition sources. Dispose of this material and its container in accordance with local regulations.
- PBS tablets (GoldBio, cat. no. P-271)
- Tris-EDTA (TE) buffer, pH 8.0, DNase-free (Sigma-Aldrich, cat. no. 93283) **! CAUTION** TE buffer is a possible irritant; use personal protective equipment including protective gloves, protective clothing and eye or face protection. Dispose of this material and its container in accordance with local regulations.
- Nuclease-free H_2O (Thermo Fisher Scientific, cat. no. AM9915G)

- 0.5 M EDTA solution (Millipore, cat. no. 324506) **! CAUTION** EDTA solution is an eye irritant; use personal protective equipment including protective gloves, protective clothing and eye or face protection. Dispose of this material and its container in accordance with local regulations.
- Tris-HCl (Sigma-Aldrich, cat. no. 10812846001) **! CAUTION** Tris-HCl is a possible irritant; use personal protective equipment including protective gloves, protective clothing and eye or face protection. Dispose of this material and its container in accordance with local regulations.
- NaCl (Sigma-Aldrich, cat. no. S9888) **! CAUTION** This product is an eye irritant and can be harmful if swallowed; use personal protective equipment including protective gloves, protective clothing and eye or face protection.
- Tryptone (Gibco, cat. no. 211705)
- Yeast extract (Gibco, cat. no. 211929)
- Agar, bacteriological (Sigma-Aldrich, cat. no. A5306)
- Milli-Q H_2O
- EtOH, analytical grade (Sigma-Aldrich, cat. no. E7023) **! CAUTION** EtOH is an eye irritant and is highly flammable; use personal protective equipment including protective gloves, protective clothing and eye or face protection. Use in a well-ventilated area and avoid ignition sources.
- Qubit dsDNA (double-stranded DNA) HS (high sensitivity) assay kit (Thermo Fisher Scientific, cat. no. Q32851) **! CAUTION** Some components are flammable and may cause irritation; use personal protective equipment including protective gloves, protective clothing and eye or face protection. Avoid breathing mist/vapors/spray and use in a well-ventilated area. Dispose of kit components in accordance with local regulations.

Transposon mutagenesis

- Glycerol (Sigma-Aldrich, cat. no G5516) **! CAUTION** Glycerol is an eye, skin and respiratory irritant; use personal protective equipment including protective gloves, protective clothing and eye or face protection. Avoid breathing mist/vapors/spray and use in a well-ventilated area. Dispose of this material and its container in accordance with local regulations.

FACS

- Isopropyl-β-D-thiogalactoside (IPTG, 100 mM in Milli-Q H_2O, filter-sterilized; Sigma-Aldrich, cat. no. I5502)

DNA extraction and quantification

- DNeasy Blood & Tissue Kit (Qiagen, cat. no. 69504) **! CAUTION** Some components are harmful if swallowed and may cause eye, skin or respiratory irritation; use personal protective equipment including protective gloves, protective clothing and eye or face protection. Avoid breathing mist/vapors/spray and use in a well-ventilated area. Dispose of kit components in accordance with local regulations.

TIS library preparation

- NEBNext Ultra II FS DNA library prep kit for Illumina (New England Biolabs, cat. no. E6177)
- NEBNext Multiplex oligos for Illumina (Index Primers Set 1) (New England Biolabs, cat. no. E7735)
- NEBNext Ultra II Q5 master mix (New England Biolabs, cat. no. M0544)
- SPRIselect (Beckman Coulter, cat. no. B23317)
- Dynabeads M-270 streptavidin beads (Invitrogen, cat. no. 65305)
- PCR enrichment primer 1 (user supplied; Table 1)
- Adaptor-specific primer (user supplied; Table 1)
- PCR enrichment primer 2 (user supplied; Table 1)

TIS library QC

- Agilent High Sensitivity DNA kit (Agilent, cat. no. 5067-4626) **! CAUTION** Kit components contain dimethyl sulfoxide (a potential mutagen). Use personal protective equipment including protective gloves, protective clothing and eye or face protection. Avoid breathing mist/vapors/spray and use in a well-ventilated area. Dispose of kit components in accordance with local regulations.
- KAPA SYBR FAST Universal qPCR kit (Roche, cat. no. KK4824) **! CAUTION** Certain kit components are harmful if swallowed and can cause eye, skin and respiratory irritation. Use personal protective equipment including protective gloves, protective clothing and eye or face protection. Avoid breathing

Table 2 | Primers required for TIS library qPCR

Name	Binding site	Sequence (5′–3′)
qPCR primer 1	Binds the Illumina P5 site	AATGATACGGCGACCACCGAG
qPCR primer 2	Binds the Illumina P7 site	CAAGCAGAAGACGGCATACGA
Transposon sequencing primer[a]	Binds the transposon	CAGGCATGCAAGCTTCAGGGTTGAGATGTG

[a]This primer is identical to the primer listed in Table 1 and will be used during Illumina sequencing.

mist/vapors/spray and use in a well-ventilated area. Dispose of kit components in accordance with local regulations.
- KAPA dilution control (Roche, cat. no. KK4906)
- qPCR primers (user supplied; Table 2)

TIS
- MiSeq Reagent Kit v3 (150 cycle; Illumina, cat. no. MS-102-3001) **! CAUTION** Certain kit components may cause skin irritation. Use personal protective equipment including protective gloves, protective clothing and eye or face protection.
- PhiX Control v3 (Illumina, cat. no. FC-110-3001)
- Transposon sequencing primer (user supplied; Table 1)
- Read 1 primer (user supplied; Table 1)

Equipment
- 1.5-ml microcentrifuge tubes
- 2.0-ml microcentrifuge tubes
- 1.5-ml DNA LoBind microcentrifuge tubes (Eppendorf, cat. no. 0030108051)
- 0.2-ml PCR tubes, thin wall (Eppendorf, cat. no. 951010006)
- Qubit assay tubes (Thermo Fisher Scientific, cat. no. Q32856)
- Bacterial culture tubes
- Glass bottles
- Erlenmeyer flasks, baffled, 150 ml and 2 liters
- Syringe Leur slip, 50 ml (BD, cat. no. 300866)
- Syringe filters, 0.22 µm (Ahlstrom-Munksjö, cat no. 760502)
- Membrane filters, 0.22 µm, 47 mm (Millipore, cat. no. GSWP04700)
- Petri dishes, 100 mm × 20 mm
- 96-well plates, polystyrene, 360-µl well volume
- qPCR plates, 96-well, fast optical, 0.1 ml (Thermo Fisher Scientific, cat. no. 4346907)
- Optical adhesive film (Thermo Fisher Scientific, cat. no. 4311971)
- Micro-cuvettes for spectrophotometry, polystyrene, 2 ml
- Multi-channel pipette, 8-channel, volume: 10–100 µl
- Falcon tubes, polypropylene, conical, 15 and 50 ml
- FACS tubes, polypropylene, round bottom, sterile, 5 ml (Corning, cat. no. 352063)
- Cryotubes, internal thread, sterile, 2 ml
- Filter pipette tips, sterile, volumes: 1–1,000 µl
- Pipette tips, sterile, volumes: 1–1,000 µl
- Vortex mixer
- Thermocycler with heated lid
- Spectrophotometer
- Incubator (static) for bacterial plates
- Incubator (shaking) for liquid bacterial cultures
- Heat block, digital, 1.5–2.0-ml microcentrifuge tube capacity
- Benchtop microcentrifuge and centrifuge
- Magnetic separation rack (New England Biolabs, cat. no. S1509S or equivalent)
- BD FACSAria Fusion or equivalent cell sorter, equipped with lasers and filters required to detect fluorophores of interest (BD Biosciences, cat. no. 656700 or equivalent)

- Qubit fluorometer (Thermo Fisher Scientific, cat. no. Q33238)
- Real-time PCR system (Thermo Fisher Scientific, cat. no. 4453535 or equivalent)
- Agilent 2100 Bioanalyzer instrument (Agilent, cat. no. G2939BA)
- MiSeq system (Illumina, cat. no. SY-410-1003)

Software

- FastQC[60] for sequencing data QC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/)
- Trimmomatic[61] (version 0.36) for read trimming, specifically removing adaptor contamination (http://www.usadellab.org/cms/?page=trimmomatic)
- Bio-TraDIS[10] package and dependencies including BWA, SMALT, Samtools, tabix and Bioconductor, for identifying transposon tag sequences, mapping reads and generating alignment files (https://sanger-pathogens.github.io/Bio-Tradis/)
- bedtools[62] for converting BAM files to BED files (https://bedtools.readthedocs.io/en/latest/)
- R[63] (version 4.0.3 or higher) (https://www.r-project.org/)
- The SorTnSeq_format_features.R script for generating a genomic feature table for the organism of interest (https://github.com/JacksonLab/SorTn-seq)[64]
- The SorTnSeq_insertion_counts.R script for processing BED files and generating a table of insertion counts (per feature) (https://github.com/JacksonLab/SorTn-seq)[64]
- The SorTnSeq_analysis.R script for analyzing the insertion counts data and identifying significantly enriched features in the different expression bins (https://github.com/JacksonLab/SorTn-seq)[64]
- R packages, including edgeR[15] for differential analysis of transposon insertion sites (https://bioconductor.org/packages/release/bioc/html/edgeR.html), as well as readxl, writexl, tidyverse, scales, ggplot2, ggrepel, ggiraph and dependencies
- (Optional) RStudio[65], an integrated development environment for R (https://rstudio.com/products/rstudio/download/)
- (Optional) PRINSEQ[66] (lite version) for read filtering or trimming (http://prinseq.sourceforge.net/)
- (Optional) Artemis[67] for sequence data visualization (https://www.sanger.ac.uk/science/tools/artemis)
- (Optional) FlowJo software (BD) for the analysis of flow cytometry data (https://www.flowjo.com/)

Files and datasets

- FASTQ sequencing files. Files from Smith et al.[13] have been deposited in the National Center for Biotechnology Information Sequence Read Archive under BioProject accession number PRJNA601789.
- Reference Sequence (RefSeq) general feature format (GFF) for the organism of interest (https://www.ncbi.nlm.nih.gov/assembly)
- FASTA nucleic acid file (.fna) for the organism of interest (https://www.ncbi.nlm.nih.gov/assembly)

Reagent setup

Lysogeny broth (LB)

Add 10 g/liter tryptone, 5 g/liter yeast extract and 5 g/liter NaCl to a glass bottle; add Milli-Q H_2O to the desired final volume and autoclave-sterilize. Store at room temperature (18–24 °C) for ≤1 month.

LB agar (1.5% (wt/vol) agar)

Add 10 g/liter tryptone, 5 g/liter yeast extract, 5 g/liter NaCl and 15 g/liter agar to a glass bottle; add Milli-Q H_2O to the desired final volume and autoclave-sterilize. Prepare fresh before each use.

PBS (100 ml)

Dissolve 1 PBS tablet in 100 ml of Milli-Q H_2O in a 200-ml glass bottle; filter (0.22 μm)- or autoclave-sterilize. Store at room temperature for ≤1 year.

0.1× TE buffer (1 ml)

Mix 100 μl of TE buffer (pH 8) with 900 μl of nuclease-free H_2O in a sterile, 1.5-ml microcentrifuge tube. Prepare fresh immediately before each use.

50% (vol/vol) glycerol (100 ml)

Mix 50 ml of glycerol and 50 ml of Milli-Q H_2O in a 200-ml glass bottle; autoclave-sterilize. Store at room temperature for ≤1 year.

80% (vol/vol) EtOH (10 ml)
Mix 8 ml of 100% ethanol and 2 ml of sterile Milli-Q H_2O in a 15-ml Falcon tube. ▲ **CRITICAL** Prepare fresh immediately before each use.

1 M Tris-HCl (pH 7.5, 100 ml)
Dissolve 15.76 g of Tris-HCl in 50 ml of Milli-Q H_2O. Adjust the pH to 7.5 by using HCl and fill up to 100 ml with Milli-Q H_2O; filter (0.22 μm)- or autoclave-sterilize. Store in a glass bottle at room temperature for ≤1 year.

1 M NaCl (100 ml)
Dissolve 5.844 g of NaCl in 50 ml of Milli-Q H_2O. Fill up to 100 ml with Milli-Q H_2O; filter (0.22 μm)- or autoclave-sterilize. Store in a glass bottle at room temperature for ≤1 year.

2× binding and washing buffer (100 ml)
The final composition is 10 mM Tris-HCl (pH 7.5), 1 mM EDTA and 2 M NaCl. Add 1 ml of 1 M Tris-HCl solution (pH 7.5), 2 ml of 1 M NaCl solution and 200 μl of 0.5 M EDTA solution to 50 ml of Milli-Q H_2O. Fill up to 100 ml with Milli-Q H_2O; filter (0.22 μm)- or autoclave-sterilize. Store in a glass bottle at room temperature for ≤1 year.

1× binding and washing buffer (100 ml)
Mix 50 ml of 2× binding and washing buffer and 50 ml of sterile Milli-Q H_2O in a 100-ml glass bottle. Store at room temperature for ≤1 year.

Adaptor dilution buffer (100 ml)
The final composition is 10 mM Tris-HCl (pH 7.5) and 10 mM NaCl. Add 1 ml of 1 M Tris-HCl solution (pH 7.5) and 1 ml of 1 M NaCl solution to 50 ml of Milli-Q H_2O. Fill up to 100 ml with Milli-Q H_2O; filter (0.22 μm)- or autoclave-sterilize. Store in a glass bottle at room temperature for ≤1 year.

Library dilution buffer (100 ml)
The final composition is 10 mM Tris-HCl, pH 8.0. Add 1 ml of 1 M Tris-HCl solution (pH 7.5) to 50 ml of Milli-Q H_2O. Adjust the pH to 8.0 by using HCl and fill up to 100 ml with Milli-Q H_2O; filter (0.22 μm)- or autoclave-sterilize. Store in a glass bottle at room temperature for ≤1 year.

50 mg/ml Km (20 ml)
Dissolve 1 g of Km in 20 ml of Milli-Q H_2O; filter (0.22 μm)-sterilize. Store at 4 °C for ≤3 months. Protect from light.

25 mg/ml Cm (20 ml)
Dissolve 0.5 g of Cm in 20 ml of 100% EtOH; filter (0.22 μm)-sterilize. Store at 4 °C for ≤3 months.

10 mg/ml Tc (20 ml)
Dissolve 0.2 g of Tc in 20 ml of 100% EtOH; filter (0.22 μm)-sterilize. Store at 4 °C for ≤3 months. Protect from light.

50 mg/ml Sm (20 ml)
Dissolve 1 g of Sm in 20 ml of Milli-Q H_2O; filter (0.22 μm)-sterilize. Store at 4 °C for ≤3 months.

50 mg/ml ALA (20 ml)
Dissolve 1 g of ALA in 20 ml of Milli-Q H_2O; filter (0.22 μm)-sterilize. Store at 4 °C for ≤3 months.

100 mM IPTG (10 ml)
Dissolve 0.238 g of IPTG in 10 ml of Milli-Q H_2O; filter (0.22 μm)-sterilize. Store 1-ml aliquots at −20 °C for ≤1 year.

Equipment setup

BD FACSAria Fusion

The laser and detector configuration of the BD FACSAria Fusion cell sorter used during *csm* SorTn-seq[13] is outlined in Supplementary Table 8. The sorter was fitted with a 70-µm nozzle (70 p.s.i.) and was operated in accordance with standard procedures. This protocol assumes familiarity with both instrument startup and sorting procedures, as well as access to standard reagents required for routine operation and QC. In advance of cell sorting, the BD FACSAria Fusion fluidics systems were initialized, and instrument QC checks were performed by using the Cytometer Setup and Tracking feature of the FACSDiva software interface. Drop delay was optimized before cell sorting.

Procedure

Transposon mutant library generation ● Timing 2 d, 3 h hands-on

▲ **CRITICAL** An overview of the mutagenesis protocol is outlined in Fig. 5. Donor and recipient strains should be inoculated from single colonies grown on LB agar with the appropriate supplements. Membrane filters should be autoclave-sterilized before use. All antibiotics and supplements are used at a final concentration of 1/1,000 of the stock solution.

1 Inoculate the donor (*E. coli* ST18 + pKRCPN2) and recipient (PCF396 + pPF1334) in 5 ml of LB with selection (ALA and Tc for the donor; Cm for recipient) in 20-ml bacterial culture tubes. Grow overnight (~16 h) at 37 °C (donor) or 30 °C (recipient) with shaking at 180 rpm.

2 After growth (16 h), remove 1.5 ml of each culture to individual 2-ml microcentrifuge tubes and centrifuge at 4,000*g* for 1 min at room temperature. Decant the supernatant, wash once with 1.5 ml of LB and then resuspend in 1.5 ml of LB.

3 Dilute washed cells (100 µl cells/900 µl of LB) in a spectrophotometric micro-cuvette and record the absorbance at 600 nm (OD_{600}).

4 Adjust the washed cells to an OD_{600} of 1 in 2 ml of LB.

5 Place 32 sterile membrane filters atop LB agar plates supplemented with ALA.

6 Pipette a 100-µl spot of the donor and recipient strains onto separate filters. These serve as no-conjugation controls.

7 Generate the conjugation mixture by adding 1.6 ml of the donor and 1.6 ml of the recipient (from Step 4) to a 15-ml Falcon tube and mix by pipetting.

8 Pipette 100 µl of the conjugation mixture to each of the 30 remaining membrane filters and allow the liquid to dry.

9 Incubate the plates (inverted) at 30 °C for 6 h without shaking.
▲ **CRITICAL STEP** 6 h was chosen for conjugation to minimize outgrowth of the library, which facilitates estimation of the starting library diversity. Conjugation times may require optimization depending on the organisms used.

10 To calculate pre-conjugation c.f.u., perform a 1/10th dilution of the donor and recipient strains from Step 4 in PBS (10 µl of washed cells in 90 µl of PBS) in a 96-well plate. Perform further 1/10th serial dilutions (to ~10^{-7}) in PBS. These dilutions should be performed in triplicate for both the donor and recipient.
▲ **CRITICAL STEP** We highly recommend calculating c.f.u. counts at each step indicated, to assess library composition, transposon mutagenesis frequency and total mutants in your final pool. See Supplementary Table 3 for *Serratia* and *E. coli* pre-conjugation c.f.u. data.

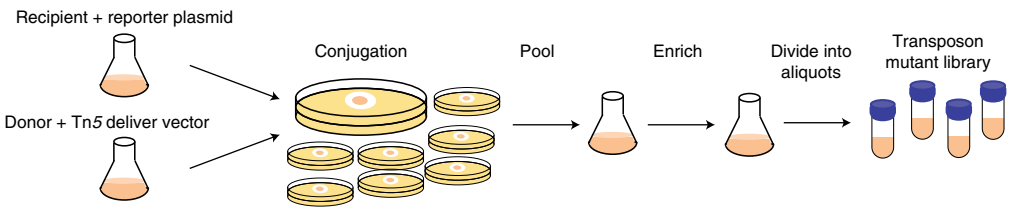


Fig. 5 | Transposon mutagenesis workflow. Cultures of the *E. coli* Tn*5* delivery vector and *Serratia* recipient are mixed in equal ratios and spotted onto filters for conjugation. After conjugation, cells are pooled and grown with selection to enrich for *Serratia* transposon mutants. The mutants are then distributed in aliquots to cryotubes for storage at −80 °C until cell sorting.

11 Using a multichannel pipette, plate spot replicates of the dilution series (10 µl) onto LB agar plates supplemented with ALA, Tc and Sm for the donor or Cm for the recipient.
12 Incubate donor plates at 37 °C and recipient plates at 30 °C (overnight, or until colonies are countable). Record the countable colony number and dilution factor for c.f.u. calculations.
13 After 6 h, aseptically transfer all 30 conjugation filters from Step 9 to one 50-ml Falcon tube and resuspend in 15 ml of LB. Vortex to remove the cells from the filters and measure the OD_{600} as described in Step 3. This is the pre-selection library.
14 Remove 50 µl of the pre-selection library to a 1.5-ml microcentrifuge tube. This will be used to estimate c.f.u.
15 Use the remaining pre-selection library to inoculate three (2-liter) baffled flasks, each containing 500 ml of LB (supplemented with Cm and Km) to a starting OD_{600} of 0.02.
16 Incubate the flasks at 30 °C for 24 h with shaking at 180 rpm. This step enriches for transposon mutants from recipients. Continue with the remaining protocol steps.
17 Place the donor and recipient no-conjugation control filters from Step 6 into separate 2-ml microcentrifuge tubes and add 1 ml of LB. Vortex to dislodge and resuspend the cells.
18 Subculture 1 ml of each resuspended control into separate 150-ml baffled flasks, each containing 25 ml of LB (supplemented with Cm and Km). Incubate the recipient control at 30 °C and the donor control at 37 °C for 24 h with shaking at 180 rpm. No growth should occur.
19 Perform serial dilutions of the pre-selection library from Step 14 in 96-well plates as described in Step 10. Plate dilutions on LB agar with the following supplements and incubate at the indicated temperature (overnight or until colonies can be enumerated):
- ALA + Sm + Tc (selects for donors; 37 °C)
- Cm (selects for recipients; 30 °C)
- Km (selects for transposon mutants; 30 °C)
- Km + Cm (selects for transposon mutants harboring the reporter plasmid; 30 °C)

See Supplementary Table 4 for pre-selection library c.f.u. data, including an estimate of the total number of transposon mutants.
20 After 24 h, transfer 15 ml of cells from each flask in Step 16 (total = 45 ml) to a 50-ml Falcon tube. Gently vortex and record the OD_{600} as previously described. This is the post-selection library.
21 Remove 1 ml of the post-selection library to a 1.5-ml microcentrifuge tube. This will be used to calculate c.f.u. (as described in Step 19) and can also be used for DNA extraction if desired. See Supplementary Table 5 for post-selection library c.f.u. data.
22 Centrifuge the remaining cells (44 ml) at 3,000*g* for 15 min at room temperature. Decant the supernatant.
23 Resuspend the cells in LB to an approximate OD_{600} of 3. For example, a 44-ml post-selection library with an OD_{600} of 1 (before centrifugation) should be resuspended in 14.6 ml of LB. This is the final transposon library.
24 Prepare 1-ml aliquots of the transposon mutant library (final OD_{600} of ~1.5) for freezer storage. Add an equal volume of 50% (vol/vol) glycerol to the final transposon library (Step 23) and mix by pipetting. Distribute 1-ml aliquots to 2-ml cryotubes until the entire transposon mutant library has been distributed in aliquots and store cryotubes at −80 °C. This should generate ≥10 transposon mutant library aliquots. If more aliquots are required, a larger volume of post-selection library can be prepared and centrifuged (Steps 20–23).

▲ CRITICAL STEP FACS, DNA extraction and TIS library construction are performed in triplicate. A minimum of three transposon mutant library aliquots are required, but preparing additional aliquots is highly recommended.

■ PAUSE POINT Transposon mutant library aliquots are stored at −80 °C until use. Typically, these libraries will be stable for multiple years; however, this may differ between strains.

Transposon mutant library outgrowth and sorting preparation ● Timing 17 h, 1.5 h hands-on

▲ CRITICAL All antibiotics and supplements are used at a final concentration of 1/1,000th of the stock solution. Control strains should be inoculated from single colonies grown on LB agar with the appropriate supplements.

25 Subculture a 1-ml aliquot of the transposon mutant library (Step 24) into 29 ml of LB (supplemented with Cm for reporter selection, Km for transposon selection and IPTG for the 2nd fluorophore (mCherry) induction) in a 150-ml baffled flask.

26 Incubate for 16 h at 30 °C with shaking at 180 rpm.
▲ CRITICAL STEP A 30-min pre-incubation without selection may be beneficial for strains carrying highly expressed reporters.

27 Inoculate control strains into 5 ml of LB (supplemented with antibiotics as outlined below) in 20-ml bacterial culture tubes and incubate for 16 h at 30 °C with shaking at 180 rpm. These cultures will be used for establishing a spectral compensation matrix during cell sorting.
- PCF396 ('unstained' control; no selection)
- PCF396 + pPF1307 ('eYFP single stain' control; Cm)
- PCF396 + pPF1439 ('mCherry single stain' control; Cm and IPTG)

28 Just before the end of the 16-h outgrowth period, prepare sterile FACS tubes for cell sorting by filling with 250 µl of PBS. This liquid helps 'cushion' cells when they are sorted into the tubes. Prepare 2 tubes for low, 2 tubes for high and 10 tubes for depleted bins and label accordingly.
▲ CRITICAL STEP Volume recovered from cell sorting can vary depending on cell concentration and sorting efficiency; therefore, the number of FACS tubes required may vary.

29 After incubation (after 16 h), record the OD_{600} of the transposon mutant library outgrowth as described in Step 3.

30 Pipette 1 ml of the outgrowth to a 1.5-ml microcentrifuge tube and spin at 4,000*g* for 1 min at room temperature. Decant the supernatant and store the tube containing the cell pellet on ice. DNA will be extracted from these cells (alongside DNA from sorted cells) and will serve as the input library.
▲ CRITICAL STEP The input libraries are not required for SorTn-seq data analysis, because the high and low expression libraries are compared against the depleted control. However, sequencing input libraries can provide information on total library diversity and gene essentiality.

31 Pipette 1.5 ml of the outgrowth to a 2.0-ml microcentrifuge tube and spin at 4,000*g* for 1 min at room temperature. Decant the supernatant, wash the pellet with 1.5 ml of PBS and then resuspend in 1.5 ml of PBS.

32 Dilute 1 ml of resuspended cells into 13 ml of PBS in a sterile 15-ml Falcon tube.
▲ CRITICAL STEP These are the cells for sorting. During cell sorting, if the sample is too dilute, add the remaining 500 µl of cells from Step 31. Likewise, if the sample is too concentrated, dilute further with additional PBS.

33 Pipette 2 µl of each control strain from Step 27 into separate FACS tubes, each containing 1 ml of PBS, and label accordingly (three tubes in total: 'unstained' control, 'eYFP single stain' control and 'mCherry single stain' control).

FACS ● Timing 3–4 h, 1–2 h hands-on

34 Set up of the cell sorter (initialization, QC and drop delay) should be performed before sample preparation (Step 32). Using the BD FACSDiva software interface, set the machine for a three-tube sort by using 'purity' mode. Apply an 'and' threshold at 400 V for both forward scatter (FSC) and side scatter (SSC).
▲ CRITICAL STEP eYFP is detected by using the 488-nm laser and 530/30 filter, whereas mCherry is detected by using the 561-nm laser and 610/20 filter. Adjust the voltage of the channels as necessary to ensure adequate detection.

35 Generate a worksheet (global sheet) containing a series of population gates (based on FSC, SSC and fluorescence parameters) as outlined in Fig. 6a.
▲ CRITICAL STEP The gate P4 should encompass mCherry-positive cells while stretching along the entire positive eYFP axis (>0). This ensures that transposon mutants with very low or high fluorescence are captured.

36 Set up and apply a compensation matrix from gate P3 by using the three control tubes ('unstained', 'eYFP single stain' and 'mCherry single stain').

37 After compensation setup, place the 15-ml Falcon tube containing the transposon mutant pool on the cell sorter. Acquire and record 100,000 events and use these data to adjust the positions of gates P1–P4 as necessary.

38 Generate a histogram of eYFP fluorescence for population P4 as indicated in Fig. 6b. Set three sort gates: low (bottom 5% of the distribution), depleted (middle 90% of the distribution) and high (top 5% of the distribution).
▲ CRITICAL STEP Because different mutants are being analyzed, the fluorescence distribution will shift slightly throughout event acquisition/cell sorting. Therefore, the percentage of cells falling within each gate is approximate. In addition, a slight separation may be needed between neighboring sort gates, because overlapping gates will cause events falling in those regions to be aborted.

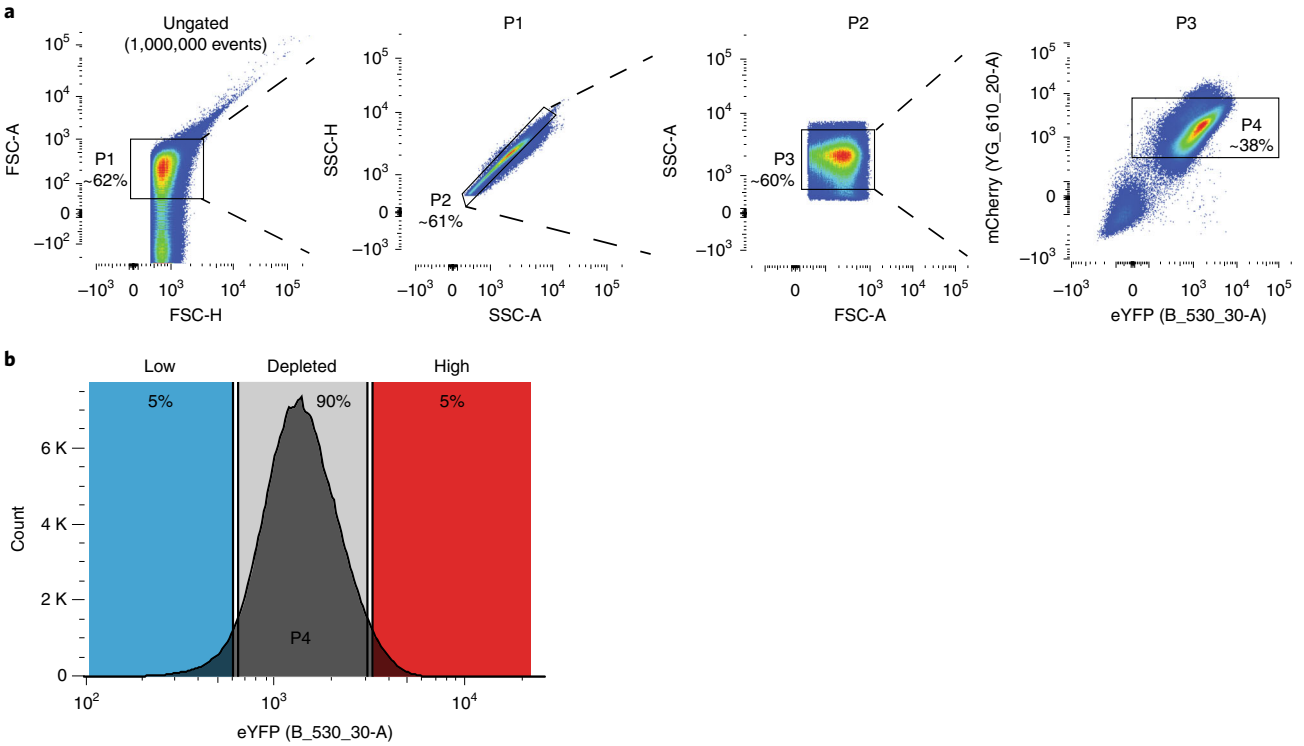


Fig. 6 | Gating strategy used during FACS. a, A series of nested gates (P1–P3) based on SSC and FSC area (-A) and height (-H) are used to isolate bacterial cell populations. The subpopulation of cells exhibiting mCherry fluorescence (YG_610_20-A; gate P4) is selected for cell sorting. PMT voltages were adjusted such that nonfluorescent cells ('unstained' control) generated a median fluorescence intensity of approximately zero. We generally observed a 2–3-log separation between fluorescent and nonfluorescent populations, which aided in delineating positive and negative populations. Data shown are from acquisition of 1,000,000 events during a representative sort, and the percentage of cells falling within each gate is indicated. **b**, Cells are sorted from population P4 on the basis of eYFP fluorescence (B_530_30-A). Sorting gates are positioned to encompass low-expressing cells (low; ~5% of P4), intermediate-expressing cells (depleted; ~90% of P4) and high-expressing cells (high; ~5% of P4). The mean ± s.d. numbers of cells per sort bin were as follows: low ($1.15 \times 10^6 \pm 4.07 \times 10^4$; mean% = 4.91), high ($9.81 \times 10^5 \pm 4.12 \times 10^4$; mean% = 4.20) and depleted ($2.12 \times 10^7 \pm 1.43 \times 10^6$; mean% = 90.88). Sorted cell numbers for each replicate can be found in the accompanying source data file. Cell sorting was performed by using a FACSAria Fusion (BD Biosciences), and data were collected with the BD FACSDiva software (version 8; BD Biosciences). Subsequent visualization was performed by using FlowJo software (version 10; BD). **a** is modified from ref. [13]. K, thousand.

39 Once gates are defined, place the labeled FACS collection tubes in the appropriate position on the machine. Initiate cell sorting and adjust the flow rate to achieve a sorting efficiency of ~70–80% (this can vary). Continue sorting until ~1×10^6 cells are sorted for the low and high gates and ~2×10^7 cells are sorted for the depleted gate.

▲ **CRITICAL STEP** We recommend sorting a minimum of 1×10^6 cells into the low and high bins to ensure adequate DNA recovery. By sorting ~2.2×10^7 cells (across all bins), the number of sorted cells was in excess (~80 fold) of our total library diversity (2.93×10^6 unique mutants). The total number of sorted cells can be scaled to increase DNA recovery or mutant library coverage.

▲ **CRITICAL STEP** Replace FACS tubes as needed throughout sorting to capture all liquid, and cap tubes when they are full to reduce contamination.

40 After completion, record the number of cells sorted into each gate (bin): low, high and depleted. Event data recorded during sorting setup can be exported as flow cytometry standard (FCS) files and analyzed in FlowJo (optional).

Cell collection ● Timing 40 min, 40 min hands-on

▲ **CRITICAL** The number of cells collected during sorting is several orders of magnitude lower than what is typically used for genomic DNA (gDNA) extraction, but within the range needed for sequencing library preparation (100 pg–500 ng). For this reason, extreme caution must be exercised when handling samples to recover adequate DNA for library preparation.

41 Pipette the contents of each depleted FACS tube into one 50-ml Falcon tube. Rinse residual cells from FACS tubes by pipetting 500 µl of PBS to each FACS tube and then transferring the liquid to the same Falcon tube.

42 Centrifuge the cells at 3,000*g* for 15 min at room temperature.

43 Gently decant the supernatant and invert the tube on a paper towel for 10 s to remove residual supernatant. A pellet will not be visible, so avoid pipetting and take care not to disturb the tip of the conical-bottom Falcon tube (where the cell pellet should be).

44 Because of smaller volumes, cells from the low and high FACS tubes can be collected by using a microcentrifuge. Pipette 1.5 ml of low or high sorted cells into 2-ml microcentrifuge tubes labeled appropriately.

45 Centrifuge the cells at 17,000*g* for 2 min.

▲ **CRITICAL STEP** Gently decant the supernatant. A pellet will not be visible, so avoid pipetting.

46 Repeat Steps 44 and 45, using the same microcentrifuge tubes, until all cells have been collected.

▲ **CRITICAL STEP** You now have cell pellets (invisible) from low, high and depleted samples, as well as the input sample (Step 30) from which DNA can be extracted. Because of the small number of cells recovered for gDNA extraction, freezing cell pellets is not recommended.

DNA extraction ● Timing 1 h, 30 min hands-on

47 Extract gDNA from bacterial cells generated in Steps 30, 43 and 46 by using the DNeasy Blood and Tissue Kit according to the manufacturer's instructions. Cell pellets should be resuspended in 180 µl of tissue lysis buffer (Buffer ATL) to begin the procedure. Elute the DNA with 33 µl of TE buffer (pH 8, 50 °C). This is the DNA for TIS library preparation.

■ **PAUSE POINT** Store DNA at 4 °C (to avoid freeze–thaw cycles and DNA shearing) until ready to proceed to quantification. We suggest storage at −20 °C for periods longer than 1 month.

DNA quantification ● Timing 20 min, 20 min hands-on

48 Quantify the DNA extracted in Step 47 by using the Qubit dsDNA HS assay kit according to the manufacturer's instructions.

49 Calculate the amount of DNA present in 26 µl for each sample. This is the starting volume used for sequencing library preparation.

▲ **CRITICAL STEP** For low and high samples, the expected DNA recovery is ~3–5 ng; for depleted samples, the expected DNA recovery is ~30 ng. We have successfully prepared sequencing libraries with as little as 3 ng of starting material.

■ **PAUSE POINT** Store DNA at 4 °C until ready to proceed to TIS library preparation.

? TROUBLESHOOTING

50 Repeat the protocol steps outlined in Steps 25–49 twice more to generate DNA ($n = 3$ for each sort bin (low, depleted and high) and unsorted control input; total $n = 12$) for transposon library sequencing preparation.

▲ **CRITICAL STEP** FACS enrichment of the transposon mutant library is performed in triplicate by using different library aliquots. Because of the cell collection and DNA extraction steps required immediately after each cell sort, we recommend performing the final two sorts (with subsequent cell collection/DNA extractions) on two different days.

TIS library preparation

▲ **CRITICAL** Sequencing libraries are prepared by using a combination of commercially available kits, custom primers, streptavidin and size-selection beads. The following procedures have been adapted from manufacturer protocols and incorporate essential manufacturer notes. For more detailed explanations, please consult the individual manufacturer protocols.

▲ **CRITICAL** Filter pipette tips should be used throughout this procedure.

DNA fragmentation and end repair ● Timing 1 h, 15 min hands-on

▲ **CRITICAL** These steps require the NEBNext Ultra II FS DNA library prep kit for Illumina.

51 Thaw the Ultra II FS reaction buffer on ice. Pipette up and down several times to break up any precipitate and vortex quickly to mix.

52 Vortex the Ultra II FS enzyme for 5 s and keep on ice.

▲ **CRITICAL STEP** The enzyme should be mixed well before use for optimal performance.

53 Add the following components to a PCR tube (on ice):

Component	Volume (µl) per one sample
DNA[a] (Step 47)	26
NEBNext Ultra II FS reaction buffer	7
NEBNext Ultra II FS enzyme mix	2
Total	35

[a]DNA should not exceed 500 ng. For input samples, prepare a dilution in 1× TE buffer by using 50 ng (total). For all other samples, use 26 µl of undiluted DNA.

54 Vortex each reaction for 5 s and briefly spin in a microcentrifuge to collect liquid.

55 Execute the following program in a thermocycler with a heated lid set to 75 °C:

Step	Time	Temperature (°C)
Fragmentation	15 min[a]	37
Inactivation	30 min	65
Hold	∞	4

[a]Fragmentation for 15 min generates DNA with a size range of 200–450 bp.

Adaptor ligation ● Timing 45 min, 15 min hands-on

▲ **CRITICAL** The following steps require the NEBNext Ultra II FS DNA library prep kit for Illumina and NEBNext Multiplex oligos for Illumina (Index Primers Set 1).

56 Prepare dilutions of the NEBNext adaptor for Illumina as required on the basis of the following table. Prepare dilutions in adaptor dilution buffer (user supplied). Routine SorTn-seq applications require a 1:10 or 1:25 adaptor dilution.

Starting DNA (total ng)	Adaptor dilution (adaptor/total)	Adaptor working concentration (µM)
100–500	No dilution	15
5–99	1:10	1.5
<5	1:25	0.6

57 Set up the ligation reaction according to the following table.

▲ **CRITICAL STEP** The NEBNext Ultra II ligation master mix is very viscous and should be mixed by pipetting up and down before use.

Component	Volume (µl) per one sample
Reaction mixture (Step 55)	35
NEBNext Ultra II ligation master mix	30
NEBNext Ultra II ligation enhancer	1
NEBNext adaptor for Illumina	2.5
Total	68.5

58 Set a 200-µl pipette to 50 µl and pipette the entire volume up and down 10 times to mix the reaction thoroughly. Briefly spin in a microcentrifuge to collect liquid.

59 Incubate samples at 20 °C for 15 min in a thermocycler with the heated lid off.

60 Add 3 µl of uracil-specific excision reagent enzyme to each sample. Mix by pipetting and incubate at 37 °C for 15 min in a thermocycler with the heated lid at ≥47 °C.

▲ **CRITICAL STEP** The uracil-specific excision reagent enzyme is required to resolve the looped structure of the NEBNext adaptor for Illumina.

Adaptor ligation cleanup ● Timing 30 min, 30 min hands-on

▲ **CRITICAL** This cleanup procedure is designed for samples with a starting DNA concentration of <100 ng. For concentrations >100 ng, consult the NEBNext Ultra II FS library prep kit for Illumina instruction manual (protocol for use with inputs ≥100 ng). For all SorTn-seq samples (including input), the following protocol was used.

61 Vortex SPRIselect beads for 30 s to resuspend.

62 Add 57 µl (0.8×) of SPRIselect to each sample and mix by vortexing samples for 5 s. Briefly spin samples in a microcentrifuge to collect liquid.
▲ **CRITICAL STEP** Centrifuge only briefly so that beads do not settle out of solution.

63 Incubate the samples at room temperature for 5 min.

64 Place the sample tubes on the magnetic separation rack to separate the beads from the supernatant.

65 After 5 min (or when the solution is clear), carefully remove and discard the supernatant by pipetting.
▲ **CRITICAL STEP** Tubes should remain on the magnet during this step.

66 While the tubes are on the magnet, add 200 µl of freshly prepared 80% (vol/vol) EtOH to each tube and incubate for 30 s.
▲ **CRITICAL STEP** Because of the hygroscopic nature of EtOH, an 80% (vol/vol) solution should be prepared fresh before each cleanup step to ensure accurate concentration.

67 While the tubes are on the magnet, carefully remove and discard the supernatant.
▲ **CRITICAL STEP** Be careful not to disturb the beads.

68 Repeat Steps 66 and 67 once for a total of two washes. Briefly spin samples in a microcentrifuge to collect residual liquid. Place samples back on the magnetic separation rack and remove all visible traces of EtOH by using a pipette tip (P10).

69 While the tubes are on the magnet, air-dry the samples with the lids open for up to 5 min.
▲ **CRITICAL STEP** Elute samples when the beads appear dark brown/glossy (with no visible liquid). If beads turn lighter brown and crack, they are over-dried, and this will result in lowered DNA recovery.

70 Remove samples from the magnetic separation rack and add 17 µl of 0.1× TE buffer to each tube. Mix by pipetting. Briefly spin samples in a microcentrifuge to collect liquid if needed.

71 Incubate samples at room temperature for 2 min.

72 Place samples back on the magnetic separation rack. After 5 min (or when the solution is clear), transfer 15 µl of supernatant to new, labeled PCR tubes.
■ **PAUSE POINT** Samples can be stored at −20 °C until continuing to PCR enrichment 1.

PCR enrichment 1 ● Timing 45 min, 15 min hands-on

▲ **CRITICAL** These steps require the NEBNext Ultra II FS DNA library prep kit for Illumina, NEBNext Multiplex oligos for Illumina (Index Primers Set 1) and user-supplied primers (Table 1).

73 Set up each PCR reaction according to the following table:

Component	Volume (µl) per one sample
Adaptor ligated DNA (Step 72)	15
NEBNext Ultra II Q5 master mix	25
PCR enrichment primer 1 (10 µM)	5
Adaptor-specific primer 1 (10 µM)	5
Total	50

74 Set a 200-µl pipette to 40 µl and mix the entire volume by pipetting up and down 10 times. Briefly spin samples in a microcentrifuge to collect liquid.

75 Perform PCR amplification in a thermocycler by using the following cycling conditions:

Step	Temperature (°C)	Time	Cycles
Initial denaturation	98	30 s	1
Denaturation	98	10 s	20[a]
Annealing/extension	65	75 s	

Table continued

(continued)

Step	Temperature (°C)	Time	Cycles
Final extension	65	5 min	1
Hold	4	∞	1

[a]For samples with a starting DNA concentration ≥50 ng, use 15 cycles; for samples ≤50 ng, use 20 cycles.

PCR enrichment 1 cleanup ● Timing 30 min, 30 min hands-on

▲ **CRITICAL** This protocol differs from the adaptor ligation cleanup (Steps 61–72) in both the concentration of SPRIselect required and elution buffer and volume. The remaining protocol steps are identical.

76 Vortex SPRIselect beads for 30 s to resuspend.

77 Add 45 µl (0.9×) of SPRIselect beads to each sample and mix by vortexing samples for 5 s. Briefly spin samples in a microcentrifuge to collect liquid.
▲ **CRITICAL STEP** Centrifuge only briefly so that beads do not settle out of solution.

78 Execute Steps 63–69.

79 Remove samples from the magnetic separation rack and add 43 µl of nuclease-free H_2O to each tube. Mix by pipetting. Briefly spin samples in a microcentrifuge to collect liquid if needed.

80 Incubate samples at room temperature for 2 min.

81 Place samples back on the magnetic separation rack. After 5 min (or when the solution is clear), transfer 40 µl of supernatant to labeled 1.5-ml DNA LoBind microcentrifuge tubes.
▲ **CRITICAL** If you wish to assess the first PCR by using Qubit or Bioanalyzer, transfer 1 µl of DNA (from the tube on the magnetic separation rack) to a PCR tube containing 4 µl of nuclease-free H_2O and store at −20 °C.

Biotinylated PCR product capture ● Timing 1 h, 30 min hands-on

▲ **CRITICAL** The following protocol has been adapted from the Dynabeads M-270 streptavidin manufacturer's protocol and incorporates essential manufacturer notes. For more detailed explanations, please consult the manufacturer's protocol.

82 Calculate the volume of streptavidin beads required for all samples by using the following formula:

$$Total\ beads\ required = 20\ \mu l\ beads \times number\ of\ samples$$

83 Vortex the beads for 30 s and remove the required volume to a 1.5-ml microcentrifuge tube. Add 1 ml of 1× binding and washing buffer.

84 Place the tube on the magnetic separation rack. After 2 min, remove the supernatant with a pipette and discard.

85 Remove the tube from the magnet and resuspend the beads in 200 µl of 1× binding and washing buffer.

86 Repeat Steps 84 and 85 twice more for a total of three washes. On the final wash, discard the supernatant.

87 Remove the tube from the magnet and resuspend the beads in 2× binding and washing buffer in a volume twice the original bead volume used in Step 83.

88 To each tube of purified PCR product (Step 81) distribute a 40-µl aliquot of washed beads. This 1:1 dilution creates the optimal salt conditions to facilitate biotin-streptavidin binding.

89 Incubate the samples at room temperature for 30 min. Flick the tubes to mix once every 5 min. Briefly spin samples in a microcentrifuge to collect liquid as needed and before proceeding to the next step.
▲ **CRITICAL STEP** Spin only briefly so that beads do not settle out of solution.

90 Place the tubes on the magnetic separation rack. After 2 min, remove the supernatant and discard.
▲ **CRITICAL STEP** Be careful not to disturb the beads, because this will result in DNA loss.

91 While the tubes are on the magnet, wash the beads with 200 µl of 1× binding and washing buffer. After 2 min, remove the supernatant.

92 Repeat Steps 90 and 91 twice more for a total of three washes. On the final wash, remove all traces of wash buffer with a pipette tip.

93 Remove the tubes from the magnet and resuspend the beads in 14 µl of nuclease-free H_2O.
94 For each sample, transfer 15 µl of resuspended beads to a labeled PCR tube.

PCR enrichment 2 ● Timing 45 min, 15 min hands-on

▲ **CRITICAL** The following steps require NEBNext Multiplex oligos for Illumina (Index Primers Set 1) and user-supplied primers (Table 1). Additional NEBNext Ultra II Q5 master mix should be ordered to facilitate this PCR enrichment step.

95 Set up each PCR reaction according to the following table:

Component	Volume (µl) per one sample
Resuspended beads (Step 94)	15
NEBNext Ultra II Q5 master mix	25
PCR enrichment primer 2 (10 µM)	5
Index primer[a]	5
Total	50

[a]A unique index primer (NEBNext Mutiplex oligos for Illumina) should be used for each sample. Take care to record which index is used for which sample, because this is required for sample identification (demultiplexing).

96 Set a 200-µl pipette to 40 µl and mix the entire volume up and down 10 times. Briefly spin samples in a microcentrifuge to collect liquid.
97 Perform PCR amplification in a thermocycler by using the following cycling conditions:

Step	Temperature (°C)	Time	Cycles
Initial denaturation	98	30 s	1
Denaturation	98	10 s	20[a]
Annealing/extension	65	75 s	
Final extension	65	5 min	1
Hold	4	∞	1

[a]For samples with a starting DNA concentration ≥50 ng, use 15 cycles; for samples ≤50 ng, use 20 cycles.

PCR enrichment 2 cleanup ● Timing 40 min, 40 min hands-on

▲ **CRITICAL** This protocol differs from the adaptor ligation cleanup (Steps 61–72) and PCR enrichment 1 cleanup (Steps 76–81) in both the concentration of SPRIselect required and elution buffer and volume. The remaining protocol steps are identical.

98 Mix each sample by pipetting and then transfer 50 µl to labeled 1.5-ml microcentrifuge tubes. Discard the PCR tubes.
▲ **CRITICAL STEP** After PCR enrichment, Dynabeads will have fallen out of solution. Ensure adequate resuspension before transferring.
99 Place the microcentrifuge tubes on the magnetic separation rack. After 3 min, transfer 47 µl of the supernatant to new, labeled 1.5-ml microcentrifuge tubes. Discard the tubes containing the beads.
100 Vortex SPRIselect beads for 30 s to resuspend.
101 Add 42.3 µl (0.9×) of SPRIselect beads to each sample and mix by vortexing samples for 5 s. Briefly spin samples in a microcentrifuge to collect liquid.
▲ **CRITICAL STEP** Centrifuge only briefly so that beads do not settle out of solution.
102 Execute Steps 63–69.
103 Remove samples from the magnetic separation rack and add 33 µl of 0.1× TE buffer to each tube. Mix well by pipetting up and down.
104 Incubate samples at room temperature for 2 min. Briefly spin samples in a microcentrifuge to collect liquid if needed.
105 Place samples back on the magnetic rack. After 5 min (or when the solution is clear), transfer 30 µl of the supernatant to labeled 1.5-ml DNA LoBind microcentrifuge tubes.
▲ **CRITICAL STEP** Be careful not to transfer any beads, because this can interfere with downstream reactions.
■ **PAUSE POINT** Store final libraries at −20 °C or proceed to Step 106.

DNA quantification ● Timing 20 min, 20 min hands-on

106 Perform DNA quantification of the sequencing libraries by using the Qubit dsDNA HS assay kit.
▲ CRITICAL STEP The expected concentration of sequencing libraries constructed from sorted cell DNA is ~10–20 ng/μl.
▲ CRITICAL STEP Filter pipette tips should be used throughout this procedure. To minimize library freeze–thaw cycles, it is recommended to perform DNA quantification and Bioanalyzer protocols in direct sequence. Alternatively, library DNA can be quantified directly after Step 105, before freezing.
? TROUBLESHOOTING

Bioanalyzer ● Timing 1.5 h, 20 min hands-on

107 Make a dilution of each library in 0.1× TE buffer to a final concentration of 2–4 ng/μl (based on the DNA concentration measured in Step 106).
▲ CRITICAL STEP Only 1 μl of DNA is loaded onto the Bioanalyzer chip, so prepare the dilution by using only a small portion of each library. The concentration does not need to be exact, because we are interested in the distribution of fragment sizes (not quantification).

108 Follow the Agilent high-sensitivity DNA assay manufacturer instructions to prepare, load and run the Bioanalyzer chip.
! CAUTION Kit components contain dimethyl sulfoxide and should be treated with the same safety precautions as all other potential mutagens.
▲ CRITICAL STEP The Bioanalyzer is used to assess fragment size distribution and quality of the sequencing libraries.

109 Determine the average fragment size of your library samples by using the region table option of the Agilent 2100 expert software (provided with the Agilent 2100 bioanalyzer and available to download for analysis-only purposes).
▲ CRITICAL STEP Calculating average fragment size by using a region table of 150–1,000 bp is generally sufficient because >80% of all fragments should be in this range. This can be adjusted as needed. A representative bioanalyzer trace of a SorTn-seq–prepared library is shown in Supplementary Fig. 2.
■ PAUSE POINT Store final libraries at −20 °C until ready to continue with sequencing library QC. DNA libraries should be stable for several years at −20 °C with limited freeze–thawing.
? TROUBLESHOOTING

qPCR ● Timing 3 h, 1.5 h hands-on

▲ CRITICAL This protocol uses the KAPA SYBR FAST Universal qPCR kit, KAPA dilution control and user-supplied qPCR primers (Table 2). Library concentrations are calculated by using a standard curve that is generated with kit components.
▲ CRITICAL Before protocol execution, prepare a worksheet detailing the position and number of all library dilutions, standards and control replicates. Perform all necessary calculations in advance to determine how much of each master mix will be required.
▲ CRITICAL The KAPA SYBR FAST qPCR master mix and carboxy-X-rhodamine (ROX) reference dyes are light sensitive. Exposure to direct light for extended periods will result in loss of fluorescent signal intensity.

110 Prepare three different dilutions of each library to be measured. Dilutions should be made fresh in 1.5-ml microcentrifuge tubes by using library dilution buffer just before qPCR. For SorTn-seq libraries (with a Qubit concentration of ~10–20 ng/μl), prepare the following dilutions: 1/10,000th, 1/100,000th and 1/200,000th. We recommend starting with a 1/100th dilution (1 μl of library in 99 μl of buffer) and then performing 1/10th serial dilutions (10 μl of library in 90 μl of buffer), followed by a final 1:2 dilution (50 μl of library in 50 μl of buffer).
▲ CRITICAL STEP A minimum of 26 μl of each library dilution is needed for qPCR reaction setup. If any dilution falls outside the range of the standard curve, it will be discarded. A minimum of two dilutions falling within the range of the standard curve are required for data analysis.

111 Prepare a 1/100,000th and 1/200,000th dilution of the dilution control, which is used to assess liquid handling accuracy.
▲ CRITICAL STEP The standards (1–6) provided with the kit are not diluted and are used ‘as is’.

112 Distribute a 20-μl aliquot of Milli-Q H_2O to a 1.5-ml microcentrifuge tube. This will serve as the non-template control.

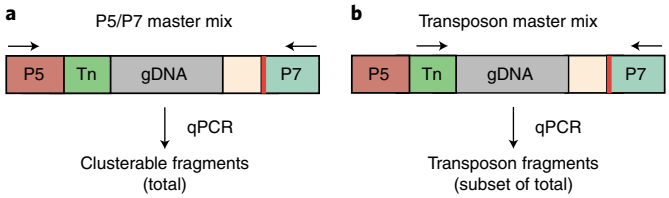


Fig. 7 | Binding sites of qPCR primers. a, The P5/P7 master mix results in amplification from the ends of the libraries (P5/P7 sites), which are required for hybridization and subsequent cluster generation on the Illumina flow cell. **b**, The transposon master mix results in amplification from the transposon and P7 end and uses the same transposon-specific primer used for Illumina sequencing. The transposon fragments should form most of the total clusterable fragments, indicating that specific products have been enriched during sequencing library preparation.

▲**CRITICAL STEP** Two different master mixes are used during qPCR (Fig. 7). The P5/P7 master mix targets the Illumina ends of library fragments. This master mix will be used to amplify each library dilution (n = no. of libraries × three dilutions), the non-template control ($n = 1$; undiluted), the dilution control ($n = 2$ dilutions) and standards 1–6 ($n = 6$; undiluted). The transposon master mix targets the transposon and the P7 Illumina end and will be used only to amplify library dilutions (n = no. of libraries × three dilutions).

113 Label 1.5-ml microcentrifuge tubes for distributing aliquots of the P5/P7 master mix. You should have one tube for each library dilution, non-template control, dilution control and standard (1–6). Label an additional set of 1.5-ml microcentrifuge tubes for distributing aliquots of the transposon master mix. You should have one tube for each library dilution.

114 Prepare the P5/P7 master mix in a 2.0-ml microcentrifuge tube according to the following table. The master mix recipe is per sample, enough for technical triplicates. Scale the volume of each component to generate enough master mix for each library dilution, dilution control, non-template control and standards 1–6.

Component	Volume (µl) per one library dilution
KAPA SYBR FAST qPCR master mix (2×)	35
qPCR primer 1 (10 µM)	1.4
qPCR primer 2 (10 µM)	1.4
Low ROX[a] (50×)	1.4
Milli-Q H_2O	16.8
Total	56[b]

[a]ROX is a passive reference dye used by qPCR machines for fluorescence normalization purposes. The appropriate concentration of ROX is indicated by machine manufacturers. [b]This volume is enough for technical triplicates of one library dilution, standard or non-template control.

115 Prepare the transposon master mix in a 2.0-ml microcentrifuge tube according to the following table. The master mix recipe is per sample, enough for technical triplicates. Scale the volume of each component to generate enough master mix for each library dilution.

Component	Volume (µl) per one library dilution
KAPA SYBR FAST qPCR master mix (2×)	35
Transposon sequencing primer (10 µM)	1.4
qPCR primer 2 (10 µM)	1.4
Low ROX (50×)	1.4
Milli-Q H_2O	16.8
Total	56

116 To the appropriate 1.5-ml microcentrifuge tubes (prepared in Step 113), distribute 52-µl aliquots of the corresponding master mix.

117 To each tube (Step 116), add 13 µl of the appropriate template (for P5/P7 tubes: library dilutions, dilution control dilutions, non-template control and standards 1–6; for transposon tubes: library dilutions). Mix by vortexing. Briefly spin samples in a microcentrifuge to collect liquid if needed.

118 Carefully pipette 20 µl into three wells on a qPCR plate (triplicate 20-µl reactions). Repeat until aliquots of all samples/dilutions have been distributed to the plate.
▲ CRITICAL STEP Pipette samples directly to the bottom of each well. If liquid collects on the sides of the wells, spin the plate in a plate centrifuge to collect liquid at the bottom.
119 Seal the qPCR plate with optical adhesive film.
120 Program the qPCR machine by using the following parameters. These settings are specific to the ViiA7 real time PCR system and may need to be adjusted depending on the machine and plate/block type/reaction speed used.

Parameter	Setting
Plate/block type	Fast 96-well
Experiment type	Standard curve
Reagent type	SYBR green
Reaction speed	Fast

121 Define the sample type and targets (dye = SYBR, quencher = NONE) and the plate layout to indicate where each sample is located.
122 Program the qPCR cycling conditions as follows and execute the run:

Step	Temperature (°C)	Time	Cycles
Initial denaturation	95	5 min	1
Denaturation	95	30 s	35
Annealing/extension/acquisition	60	45 s	
Melt curve	65–95	-	1

123 After the run, export the data as a text file and analyze by using the KAPA library quantification data analysis template for Illumina (available by request from sequencing.roche.com/support). Alternatively, data can be analyzed manually by following instructions in the technical data sheet for KAPA library quantification kits (http://technical-support.roche.com/default.aspx).
? TROUBLESHOOTING
124 Compare the average molarities of the libraries amplified with the two different master mixes. The P5/P7 master mix concentration represents all clusterable fragments (containing Illumina flow cell binding sites). The transposon master mix concentration represents all fragments containing the transposon sequencing primer binding site.
▲ CRITICAL STEP During *csm* SorTn-seq[13], 84% of clusterable fragments (P5/P7) harbored the transposon sequencing primer binding site (as averaged across libraries). After sequencing, 94% of reads (as averaged across libraries) contained the specified transposon tag (allowing up to two mismatches).
■ PAUSE POINT Store final libraries at −20 °C until ready to continue to Illumina sequencing sample preparation. DNA libraries should be stable for several years at −20 °C with limited freeze–thawing.

Illumina sequencing sample preparation ● Timing 30 min, 30 min hands-on

125 Refer to the Qubit library concentrations (Step 106) and the Bioanalyzer average fragment sizes (Step 109) to calculate the concentration (nM) of each library, according to the following formula:

$$Concentration\,(nM) = \left(\frac{Qubit\ library\ concentration(ng/\mu l)}{660\frac{g}{mol} \times Bioanalyzer\ average\ library\ size\ (bp)}\right) \times 10^6$$

126 Prepare 10-nM dilutions of each library in 10–20 µl of 0.1× TE buffer in 1.5-ml DNA LoBind microcentrifuge tubes.
127 Generate a final library pool (≥30 µl) for sequencing by pooling aliquots of the 10-nM library dilutions in a DNA LoBind 1.5-ml microcentrifuge tube. Libraries can be pooled in different ratios.
▲ CRITICAL STEP For the *csm* SorTn-seq experiment[13], libraries were pooled in the following

ratios: input ($n = 3$; each 14% of total), depleted ($n = 3$; each 10% of total), low ($n = 3$; each 4.67% of total) and high ($n = 3$; each 4.67% of total). Recently, a SorTn-seq experiment was successfully executed with libraries pooled in equal ratios (H. Hampton, personal communication). This pool did not contain input libraries.

PAUSE POINT Store the final library pool (and 10-nM dilutions) at −20 °C until ready for Illumina sequencing. DNA libraries should be stable for several years at −20 °C with limited freeze–thawing.

Illumina sequencing ● Timing 1–2 d

CRITICAL Sequencing is performed on an Illumina MiSeq by using the MiSeq reagent kit v3 150 cycle.

128 PhiX control library (10%) is added to the final library pool to increase diversity and serve as an internal control. Load custom Read 1 primers (transposon sequencing primer and Illumina Read 1 primer; Table 1) in cartridge position 18 and amend the sample sheet in accordance with standard Illumina procedure. Use the standard Illumina primer mix (cartridge position 13) for index sequencing. Prepare the library/PhiX pool for sequencing according to standard Illumina procedure and load at 1.5 pM.

CRITICAL STEP Libraries must be loaded at lower than standard concentrations because of low nucleotide diversity in the first 12 sequencing cycles (due to the transposon tag).

129 Sequence libraries (single-ended) for 75 cycles. After sequencing, FASTQ files are automatically generated as part of the Illumina MiSeq Local Run Manager standard workflow. This pipeline uses Illumina bcl2fastq software for demultiplexing and run metrics/QC, and fastqc and multiqc software for read QC.

Data analysis

CRITICAL The computational analyses used in this protocol have been designed for a Unix-like operating system with a bash shell. The first stage of the analysis (Steps 130–136) is run in a terminal window. The second stage of the analysis (Steps 139–141) can be run by using RStudio's interactive environment. To ensure accessibility and ease of use for those working with diverse organisms, the data analysis pipeline (Fig. 8) has been modified from Smith et al.[13] as follows. We have included another option for read trimming (using Trimmomatic), which can be used to trim user-specified adaptor sequences. A new script (SorTnSeq_format_features.R) has been included, which allows the user to generate a SorTn-seq compatible table ('[genome.prefix]_features_sortnseq.xlsx') that specifies the genomic features in their organism (based on the corresponding RefSeq GFF file) and adds intergenic regions as features. To assign reads to genomic features, the updated R script (SorTnSeq_insertion_counts.R) uses browser extensible data (BED) files (derived from binary alignment map (BAM) files) rather than insert site plot files (.plot). This change simplifies the analysis for genomes with multiple contigs or replicons (because the TraDIS pipeline outputs separate plot files for each contig/replicon) but results in slight differences in the mapping information provided in the plot files versus BED files, because of how soft-masked bases are resolved by TraDIS. We also updated how overlapping features are resolved, to split the reads and unique insertions equally between the corresponding features. Several functions to improve the overall usability have been added, such as a sample name and file parsing system and improved plots for data visualization. Overall, these improvements resulted in minor changes to the final output of the *Serratia csm* example dataset (Table 3 and Supplementary Data 1).

FASTQ quality control using FastQC ● Timing 30 min, 30 min hands-on

CRITICAL The following steps describe QC of sequencing FASTQ files and are performed in a terminal window of a Unix-like operating system with a bash shell. Ensure that all FASTQ files are in the appropriate directory before proceeding.

130 Process all FASTQ files on the command line by using the following command:

```
fastqc -t [THREADS] *.fastq.gz
```

This generates an analysis (.html) for each sample.

CRITICAL STEP The transposon tag sequence results in very limited diversity for the first 12 nt of the sequencing run. Thus, it is expected that libraries will fail some standard QC metrics, such as per-base sequence content, overrepresented sequences and k-mer content. This is not a cause for concern. If significant adaptor contamination is indicated, read trimming is advised. If sequence duplication levels are very high, consider reducing the number of cycles used in PCR enrichment 2.

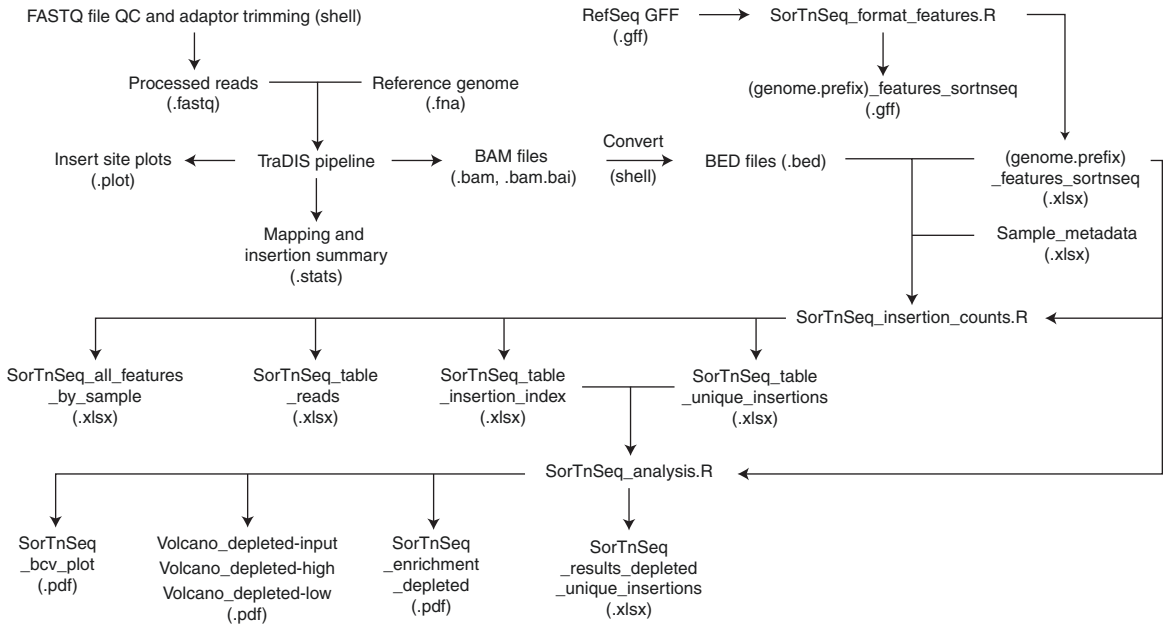


Fig. 8 | Summary of input and output files of the SorTn-seq analysis. FASTQ files are first processed to assess quality and remove adaptor contamination in the terminal window (shell). Processed files are fed into the TraDIS pipeline to identify the transposon tag and map reads to the reference genome (.fna file). The TraDIS pipeline summarizes mapping and insertion statistics (.stats file) and produces sample-specific files, such as reads per nucleotide position (.plot files) and binary alignment map (BAM) files and indices (.bam and .bam.bai). In the terminal, BAM files are converted to browser extensible data (BED) files (.bed) for subsequent analysis in R. To assign mapped reads to specific genomic features, an organism-specific feature table ([genome.prefix]_features_sortnseq.xlsx) is first generated in R (SorTnSeq_format_features.R), which parses RefSeq general feature format (GFF) files and adds intergenic regions as features. The feature table, BED files and user-supplied sample information (sample_metadata.xlsx) are used to generate tables of read counts, insertion counts and insertion index (number of insertions/feature length) for each sample (.xlsx files). To identify differentially enriched features, the unique insertion table (SorTnSeq_unique_insertions.xlsx) and insertion index table (SorTnSeq_table_insertion_index.xlsx) are processed by using edgeR (SorTnSeq_analysis.R). A table summarizing feature enrichment (SorTnSeq_results_depleted_unique_insertions.xlsx) is generated along with plots that summarize the results (.pdf) (Fig. 9).

Read trimming using Trimmomatic ● Timing 30 min, 10 min hands-on

▲ **CRITICAL** The following steps can be used to trim a specified adaptor sequence from reads and are performed in a terminal window. Reads in Smith et al.[13] were originally trimmed to 50 nt from the 3′ end by using PRINSEQ (Supplementary Protocol).

▲ **CRITICAL** Reads should not be trimmed from the 5′ end, because this will remove the transposon tag sequence that is required for read mapping using the Bio-TraDIS pipeline.

131 Trim the specified FASTQ file by using the following command code on the command line. This generates a new file that has been appropriately trimmed. Key parameters are explained in the table following the code.

```
trimmomatic SE -threads 20 -trimlog trim_summary input.fastq.gz output.fastq.gz ILLUMINACLIP:TruSeq3-SE:2:30:1
```

Parameter	Action
`SE`	Specifies the read type (SE = single ended)
`-threads`	Specifies the number of threads for the server to use (20 is standard)
`-trimlog`	Specifying a trimlog file creates a log of all read trimmings ('trim_summary')
`input.fastq.gz`	Specifies the FASTQ file name
`output.fastq.gz`	Specifies the trimmed FASTQ file name
`ILLUMINACLIP:`	Specifies to cut the adapter and other Illumina-specific sequences from the read
`TruSeq3-SE:2:30:1`	Specifies the reference Illumina adaptor sequences (FASTA)[a] and matching thresholds

[a]Provided as part of the Trimmomatic pipeline. Custom files can also be generated.

132 Repeat Step 131 as necessary until all files have been trimmed.

Table 3 | Subset of data from the differential enrichment analysis

type	start	end	product	id	low_logFC	low_p.value	low_p.adj
gene	882635	885190	class I adenylate cyclase	gene-CWC46_RS04070	2.22	2.26×10^{-39}	1.14×10^{-35}
intergenic	2489250	2490575		intergenic_NZ_CP025085.1-nt02489250	1.26	5.25×10^{-32}	1.32×10^{-28}
gene	2490932	2491510	flagellar transcriptional regulator FlhC	gene-CWC46_RS11350	1.60	3.81×10^{-29}	6.40×10^{-26}
gene	968371	969003	cAMP-activated global transcriptional regulator CRP	gene-CWC46_RS04470	3.00	2.36×10^{-21}	2.97×10^{-18}
gene	2490576	2490926	flagellar transcriptional regulator FlhD	gene-CWC46_RS11345	1.39	1.47×10^{-14}	1.48×10^{-11}
gene	2517774	2518556	flagellar type III secretion system protein FliR	gene-CWC46_RS11485	1.13	1.42×10^{-13}	1.19×10^{-10}
gene	4734270	4736231	hypothetical protein	gene-CWC46_RS21600	1.04	2.03×10^{-12}	1.46×10^{-09}
gene	4741195	4742277	CDP-glucose 4,6-dehydratase	gene-CWC46_RS21630	1.08	3.67×10^{-12}	2.31×10^{-09}
gene	2548972	2549694	RNA polymerase sigma factor FliA	gene-CWC46_RS11610	1.07	7.19×10^{-11}	4.02×10^{-08}
gene	2565506	2567074	glucan biosynthesis protein G	gene-CWC46_RS11700	1.10	1.54×10^{-10}	7.05×10^{-08}
gene	3669680	3670336	UvrY/SirA/GacA family response regulator transcription factor	gene-CWC46_RS16790	1.03	2.51×10^{-09}	1.05×10^{-06}
gene	3626499	3628619	ATP-dependent DNA helicase DinG	gene-CWC46_RS16585	1.57	6.60×10^{-09}	2.56×10^{-06}
gene	4739385	4740368	glycosyltransferase family 2 protein	gene-CWC46_RS21620	1.00	9.41×10^{-09}	3.39×10^{-06}
gene	3106069	3107085	two-component system response regulator RssB	gene-CWC46_RS13965	1.00	4.71×10^{-08}	1.58×10^{-05}
intergenic	969004	969324		intergenic_NZ_CP025085.1-nt00969004	1.70	1.46×10^{-07}	4.34×10^{-05}

This table was modified from the final output ('SorTnSeq_results_depleted_unique_insertions.xlsx'; Step 141) to show the top 15 enriched features (based on adjusted *P* value and $\log_2$ fold change >0.5) in the low expression bin. The $\log_2$ fold change (low_logFC), *P* value (low_p.value) and adjusted (i.e., adjusted in edgeR by using Benjamini-Hochberg correction to control false discovery rate) *P* value (low_p.adj.) are displayed for each feature. In the full output, the $\log_2$ fold changes, *P* value and adjusted *P* value are shown for each feature in each sort bin (low, high and input), as well as $\log_2$ counts per million, strand information and gene names (Supplementary Data 1). Source data were generated in ref. [13].

Transposon tag identification and read mapping ● Timing 1 h, 10 min hands-on

▲ **CRITICAL** This analysis is performed in a terminal window by using the Bio-TraDIS[10] pipeline. This pipeline identifies a specified transposon tag, trims it and then maps the reads to a reference file by using either Burrows-Wheeler Aligner (BWA)[68] or SMALT[69] tools.

▲ **CRITICAL** For detailed explanations ft he Bio-TraDIS pipeline[10], consult the GitHub page (https://github.com/sanger-pathogens/Bio-Tradis/).

133 Obtain the RefSeq FASTA nucleic acid (.fna) file for the organism of interest and load it to the appropriate directory.

134 Ensure that FASTQ files are in the appropriate directory and prepare a list of all FASTQ files to be analyzed:

```
dir *.fastq.gz >filelist.txt
```

135 Run the following code on the command line. Key parameters are explained in the table following the code.

```
bacteria_tradis --smalt --smalt_k 10 --smalt_s 1 --smalt_y 0.92
--smalt_r -1 -mm 2 -v -f filelist.txt -t TATAAGAGACAG -r genome.fna
```

Parameter	Action
`--smalt`	Specifies which mapping tool to use
`--smalt_k`[a]	Specifies the word length (k-mer size) used to generate the reference FASTA index. This is the minimum length of exact match required to trigger an alignment attempt (generally 10–20 for bacterial genomes)
`--smalt_s`	Specifies the step size for k-mers in the reference FASTA index A value of `'1'` indicates that every k-mer word along the reference is indexed
`--smalt_y`	Specifies the minimum identity threshold between read/reference for mapping to be reported
`--smalt_r`	Specifies what happens to reads that map equally well to multiple locations. Specifying `'-1'` indicates that these reads are discarded

Table continued

(continued)

Parameter	Action
`-mm`[b]	Specifies the number of mismatches allowed in a transposon tag sequence
`-v`	Details the status of running processes (verbose)
`-f`	Specifies the text file listing each FASTQ file to be analyzed
`-t`	Specifies the transposon tag sequence
`-r`	Specifies the reference genome in FASTA nucleic acid (.fna) format

[a]For a full description of SMALT mapping parameters, execute the command `smalt map -H` or `smalt index -H` on the command line. [b]Per-base sequencing quality (assessed through FastQC) may be lower in the first 12 nt of the sequencing run because of the low diversity of the transposon tag. For this reason, we allow up to two mismatches in the transposon tag sequence.

This generates a mapping and insertion summary file (.stats), as well as BAM files and indices (.bam and .bam.bai) and insert site plot files (.plot), which tabulate the number of reads at each nucleotide position for each sample.

▲ **CRITICAL STEP** The transposon tag sequence and reference FASTA will vary depending on the transposon and organism used.

? TROUBLESHOOTING

136 Insert site plot files can be opened in Artemis to visualize insertion sites. Convert BAM files to BED files before proceeding:

```
for FILE in *.bam; do
bedtools bamtobed -i $FILE > $FILE.bed
done
```

137 Transfer the BED files to the appropriate directory for subsequent analysis using R.

Generating a master feature table ● Timing 10 min, 10 min hands-on

▲ **CRITICAL** This analysis requires the RefSeq .gff file for the organism of interest.

138 Obtain the RefSeq .gff file for the organism of interest and update the corresponding 'genome.prefix' variable in SorTnSeq_format_features.R.

139 Run SorTnSeq_format_features.R to generate a new feature table ('[genome.prefix]_features_sortnseq.xlsx') that specifies intergenic regions as well as RefSeq features.

Compiling unique insertion counts ● Timing 20 min, 20 min hands-on

▲ **CRITICAL** This analysis requires BED files (generated during Step 136), the feature table generated in Step 139 ('[genome.prefix]_features_sortnseq.xlsx') and sample information ('sample_metadata.xlsx') detailing BED file name and corresponding library type and replicate (e.g., low 1) for each sample. Because some features can tolerate insertions at the 3′ end without disrupting function[27], we trimmed reads mapping to the final 10% (3′ end) of all non-intergenic features (see optional parameters below).

140 Run the SorTnSeq_insertion_counts.R script to generate the following files:

- 'SorTnSeq_table_reads.xlsx': summarizes the number of reads per feature for each library
- 'SorTnSeq_table_insertion_index.xlsx': summarizes the insertion index (number of insertions/feature length) per feature for each library
- 'SorTnSeq_table_unique_insertions.xlsx': summarizes the number of unique transposon insertions per feature for each library
- 'SorTnSeq_all_features_by_sample.xlsx': summarizes the number of reads, unique insertions and insertions index per feature for each library

Parameter	Action	Default
`trim.3.prime<-X`	X specifies the proportion (0–1) of the 3′ end of features (non-intergenic) to exclude for read counting	0.1
`trim.5.prime<-X`	X specifies the proportion (0–1) of the 5′ end of features (non-intergenic) to exclude for read counting	0

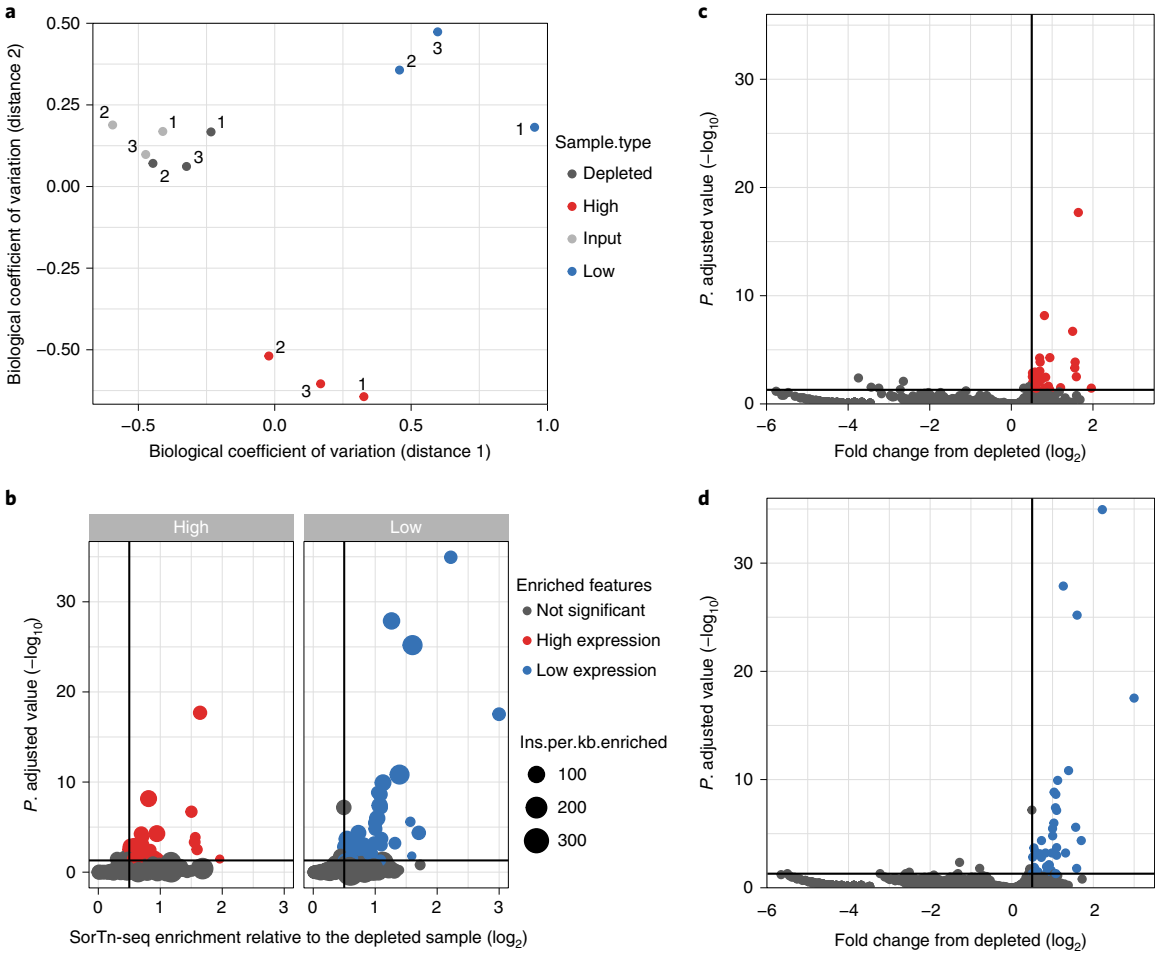


Fig. 9 | Example plots generated by the SorTnSeq_analysis R script. a, MDS plot based on biological coefficient of variation, constructed with the top 100 most-variable features. **b**, SorTn-seq enrichment plot displaying significantly enriched features (*P* adjusted < 0.05 and $\log_2$ fold change > 0.5) in both the high and low bins. The size of each point indicates enrichment of insertions in each feature, per kilobase. This plot is interactive in R, where hovering over a point displays the feature name. **c** and **d**, Volcano plots showing the adjusted *P* value ($-\log_{10}$) versus the $\log_2$ fold change of features in the high (**c**) and low (**d**) bins compared against the depleted control. Source data were generated in ref. [13].

Regulator prediction using edgeR ● **Timing 20 min, 20 min hands-on**

▲ **CRITICAL** This analysis requires '[genome.prefix]_features_sortnseq.xlsx' (generated during Step 139), 'SorTnSeq_table_unique_insertions.xlsx' and 'SorTnSeq_table_insertion_index.xlsx' (generated during Step 140).

141 Run the SorTnSeq_analysis R script to generate the following files:

- 'SorTnSeq_bcv_plot.pdf': multidimensional scaling plot (MDS) to visualize the similarity between libraries and replicates based upon the biological coefficient of variation (Fig. 9a)
- 'SorTnSeq_enrichment_depleted.pdf': summarizes the enriched features in the high and low bins, at the specified cutoff values (Fig. 9b). In R, this plot is interactive, and hovering above each point displays the feature name.
- 'volcano_depleted-low.pdf', 'volcano_depleted-high.pdf' and 'volcano_depleted-input.pdf': volcano plots for the different libraries, which show the $\log_2$ fold change and adjusted *P* value for each analyzed feature, as compared against the reference sample (Fig. 9c,d)
- 'SorTnSeq_results_depleted_unique_insertions.xlsx': results of the differential enrichment analysis. The 'classic' approach used in edgeR estimates common and tagwise (feature-wise) dispersion (Supplementary Fig. 3) by using the quantile-adjusted conditional maximum likelihood method[70,71] and performs an exact test (comparable to Fisher's exact test). The results of the exact test (including $\log_2$ fold changes against the reference sample, and adjusted

P values), as well as feature names and descriptions, are written to this file (Table 3 and Supplementary Data 1).

The following parameters are used during analysis:

Parameter	Action	Default
`bcv.features<-X`	X specifies the number of features used to model the biological coefficient of variation (BCV) for plotting purposes	100
`read.cutoff.depleted<-X`	X specifies the minimum number of insertions that must be present in all replicates of the reference sample for a feature to be analyzed for enrichment by edgeR	2
`reference.sample<-"X"`	X specifies which samples to use as the control group, either the depleted or input libraries	depleted
`threshold.fc<-X`	X specifies the $\log_2$ fold change cutoff for plots	0.5
`threshold.p.adj<-X`	X specifies the significance value (adjusted *P* value) cutoff for plots	0.05

? TROUBLESHOOTING

Troubleshooting

Troubleshooting advice can be found in Table 4. Refer to individual manufacturer protocols for advanced troubleshooting options.

Table 4 | Troubleshooting table

Step	Problem	Possible reason	Solution
49	Low DNA yield from sorted cells	Loss of cells during collection or DNA extraction	Exercise care during sorted cell recovery and DNA extraction. Ensure that FACS tubes are rinsed to remove all cells
106	Low DNA yield from final sequencing libraries	Loss of sample during cleanup procedures	Exercise care when washing/eluting beads—ensure that the supernatant is retained/discarded as indicated. Run the final library and post-PCR enrichment 1 DNA (Step 81) on the Bioanalyzer/Qubit to assess if DNA was present after PCR enrichment 1
		Inefficient enrichment of the transposon DNA	The number of PCR cycles may need to be optimized. If the final library molarity is <10 nM, transposon sequencing library preparation should be repeated
109	Unexpected peaks present in the Bioanalyzer trace	Primer or adaptor-dimer contamination	Significant primer (~80 bp) or adaptor-dimer (~128 bp) contamination should be removed by repeating the PCR enrichment 2 cleanup Steps 100–105 by using 0.9× SPRI-select beads. Adaptor-dimer contamination can also be assessed through the melt curve generated during qPCR (Step 122)
		PCR artifacts	Reduce the number of PCR enrichment 2 cycles
		gDNA contamination	Small quantities of gDNA (fragments > upper marker) should not interfere with sequencing, because they will cluster less efficiently on the flow cell
123	The standard curve does not fall within expected parameters	Early amplification of standard 1; poor liquid handling (see below)	DNA Standard 1 is highly concentrated and amplifies at a very early cycle threshold/quantification cycle (Ct/Cq). Adjust the cycles during which the machine establishes baseline fluorescence to cycles 1–3 (instead of the typical cycles 3–15)
	Library dilutions fall outside the range of the standard curve	Improper library dilution	Verify library dilutions; the range of dilutions assayed should be altered if libraries are too concentrated or dilute
	High standard deviation among library dilutions	Poor liquid handling	Take care when pipetting and dispensing samples to the qPCR plate; ensure that dilution control is used to assess liquid handling
135	Low number of reads with transposon 'tag'	Mismatch cutoff (-mm) too stringent	Assess read QC metrics by using FastQC. If samples show low quality in the transposon 'tag' region, try a less stringent cutoff

Table continued

Table 4 (continued)

Step	Problem	Possible reason	Solution
	Low number of reads mapping to reference	Mapping parameters too strict	Adjust the mapping stringency (--smalt_y)
		Large number of reads mapping to plasmids	Map reads to all other possible sources (the reporter plasmid, transposon delivery plasmid and any host plasmids)
		Library contamination	Libraries could be contaminated by other samples prepared by using the same transposon (can be verified by mapping to the other reference sequence)
141	High variability among library replicates	Variation during sorting of transposon mutant library or sequencing library preparation	Ensure consistency between transposon mutant library outgrowth, cell sorting and sequencing library preparation

Timing

Steps 1–33, transposon mutant library generation, transposon mutant library outgrowth and sorting preparation: 3 d, 4.5 h hands-on
Steps 34–50, FACS, cell collection, DNA extraction and DNA quantification: 5–6 h, 2–4 h hands-on
Steps 51–106, DNA fragmentation and end repair, adaptor ligation, adaptor ligation cleanup, PCR enrichment 1, PCR enrichment 1 cleanup, biotinylated PCR product capture, PCR enrichment 2, PCR enrichment 2 cleanup and DNA quantification: 6–7 h, 3–4 h hands-on
Steps 107–124, Bioanalyzer and qPCR: 4–5 h, 2 h hands-on
Steps 125–129, Illumina sequencing sample preparation and Illumina sequencing: 1–2 d, 30 min hands-on
Steps 130–141, FASTQ QC using FastQC, read trimming using Trimmomatic, transposon tag identification and read mapping, generating a master feature table, compiling unique insertion counts and regulator prediction using edgeR: 3 h, 1.5–2 h hands-on

Anticipated results

SorTn-seq enables the high-throughput genome-wide discovery of regulators of bacterial gene expression. The SorTnSeq_analysis R script generates a series of graphs including MDS plots and volcano plots (Fig. 9). MDS plots approximate the difference in expression (in this case, insertions) between samples on the basis of differences between different parameters, such as the BCV (Fig. 9a) or fold changes. Under the edgeR model, biological variation is considered the main source of variation arising between samples. Dispersion determines how the variance between each feature is modeled in edgeR, and thus how the mean counts of each feature are estimated. When count data (insertion counts) are fit to a negative binomial distribution, the BCV represents the square root of the dispersion. 'Common' dispersion estimates the mean dispersion across all features, whereas 'tag-wise' dispersion estimates the dispersion of each feature individually[15] (Supplementary Fig. 3a).

The 'SorTnSeq_enrichment_depleted.pdf' plot visualizes enriched features in the high or low bin, according to user-defined cutoffs for $\log_2$ fold change and adjusted *P* values (Fig. 9b). In R, this plot is interactive, and feature names are displayed when hovering over a point of interest. Volcano plots are also generated for each sort bin (Fig. 9c,d) and visualize the $\log_2$ fold change and adjusted *P* value of all analyzed features. Mean difference plots can also be generated, which show the average $\log_2$ fold change versus average $\log_2$ counts per million for each feature, for each library type (Supplementary Fig. 3b).

The culmination of the SorTn-seq experiment and data analysis results in an output table (generated in Step 141) that summarizes the fold changes ($\log_2$) of features in the different sort bins (low/high) as compared to the depleted control. Each $\log_2$ fold change is associated with an adjusted *P* value (to control for false discovery rate) to assess statistical significance. For simplicity, a subset of the data is shown in Table 3 (full data in Supplementary Data 1). During type III-A CRISPR-Cas SorTn-seq analysis[13], we identified potential regulators as features having a $\log_2$ fold change >0.5 and an adjusted *P* value < 0.05, indicating that the feature was 'enriched' in the respective sort bin (low/high). Using this data analysis pipeline, we identified 44 features enriched in the low bin (36 genes,

1 pseudogene and 7 intergenic regions) and 40 features in the high bin (34 genes, 1 pseudogene and 5 intergenic regions). To assess the robustness of our results, we confirmed expression changes in 19 different regulators identified through the *csm* SorTn-seq screen[13].

In conclusion, SorTn-seq provides the ability to undertake a nonbiased discovery on a genome-wide level of genomic features that affect the regulation of any gene of interest. SorTn-seq is applicable for uncovering genes controlling any other gene of interest in bacteria where transposon mutagenesis and fluorescent reporters are available. Therefore, the SorTn-seq method has the power to reveal regulatory processes underpinning important bacterial traits such as virulence, symbiosis or any other physiological pathway of interest.

Data availability

Sequencing data originally published in ref. [13] are available in the Sequence Read Archive under BioProject number PRJNA601789. The annotated genome of *Serratia* sp. ATCC 39006-LacA is available through the National Center for Biotechnology Information (reference sequence NZ_CP025085.1). Source data are provided with this paper.

Code availability

R scripts and files required for data processing the *Serratia csm* dataset are available at Zenodo (https://doi.org/10.5281/zenodo.4554398) and on GitHub (https://github.com/JacksonLab/SorTn-seq)[64].

References

1. Martínez-García, E. & de Lorenzo, V. Transposon-based and plasmid-based genetic tools for editing genomes of gram-negative bacteria. in *Synthetic Gene Networks: Methods and Protocols* (eds. Weber, W. & Fussenegger, M.) 267–283 (Springer, 2012).
2. Munoz-Lopez, M. & Garcia-Perez, J. L. DNA transposons: nature and applications in genomics. *Curr. Genomics* **11**, 115–128 (2010).
3. Lampe, D. J., Grant, T. E. & Robertson, H. M. Factors affecting transposition of the Himar1 mariner transposon in vitro. *Genetics* **149**, 179–187 (1998).
4. Rubin, E. J. et al. In vivo transposition of mariner-based elements in enteric bacteria and mycobacteria. *Proc. Natl Acad. Sci. USA* **96**, 1645–1650 (1999).
5. Goryshin, I. Y., Miller, J. A., Kil, Y. V., Lanzov, V. A. & Reznikoff, W. S. Tn5/IS50 target recognition. *Proc. Natl Acad. Sci. USA* **95**, 10716–10721 (1998).
6. van Opijnen, T. & Levin, H. L. Transposon insertion sequencing, a global measure of gene function. *Annu. Rev. Genet* **54**, 337–365 (2020).
7. 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).
8. Goodman, A. L. et al. Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe* **6**, 279–289 (2009).
9. Langridge, G. C. et al. Simultaneous assay of every *Salmonella* Typhi gene using one million transposon mutants. *Genome Res.* **19**, 2308–2316 (2009).
10. Barquist, L. et al. The TraDIS toolkit: sequencing and analysis for dense transposon mutant libraries. *Bioinformatics* **32**, 1109–1111 (2016).
11. Davey, L. & Valdivia, R. H. Bacterial genetics and molecular pathogenesis in the age of high throughput DNA sequencing. *Curr. Opin. Microbiol.* **54**, 59–66 (2020).
12. Cain, A. K. et al. A decade of advances in transposon-insertion sequencing. *Nat. Rev. Genet* **21**, 526–540 (2020).
13. Smith, L. M. et al. The Rcs stress response inversely controls surface and CRISPR-Cas adaptive immunity to discriminate plasmids and phages. *Nat. Microbiol.* **6**, 162–172 (2021).
14. Givan, A. L. *Flow Cytometry: First Principles* (John Wiley & Sons, 2013).
15. 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).
16. Pessi, G., Blumer, C. & Haas, D. lacZ fusions report gene expression, don't they? *Microbiology* **147**, 1993–1995 (2001).
17. Abuaita, B. H. & Withey, J. H. Genetic screening for bacterial mutants in liquid growth media by fluorescence-activated cell sorting. *J. Microbiol. Methods* **84**, 109–113 (2011).
18. Binder, S. et al. A high-throughput approach to identify genomic variants of bacterial metabolite producers at the single-cell level. *Genome Biol.* **13**, R40 (2012).
19. Hassan, K. A. et al. Fluorescence-based flow sorting in parallel with transposon insertion site sequencing identifies multidrug efflux systems in *Acinetobacter baumannii*. *mBio* **7**, e01200-16 (2016).
20. Dorman, M. J., Feltwell, T., Goulding, D. A., Parkhill, J. & Short, F. L. The capsule regulatory network of *Klebsiella pneumoniae* defined by density-TraDISort. *mBio* **9**, e01863-18 (2018).

21. Duncan, M. C. et al. High-throughput analysis of gene function in the bacterial predator *Bdellovibrio bacteriovorus*. *mBio* **10**, e01040-19 (2019).
22. Rego, E. H., Audette, R. E. & Rubin, E. J. Deletion of a mycobacterial divisome factor collapses single-cell phenotypic heterogeneity. *Nature* **546**, 153–157 (2017).
23. Fowler, C. C. & Galan, J. E. Decoding a *Salmonella* typhi regulatory network that controls typhoid toxin expression within human cells. *Cell Host Microbe* **23**, 65–76.e6 (2018).
24. Thibault, D. et al. Droplet Tn-Seq combines microfluidics with Tn-Seq for identifying complex single-cell phenotypes. *Nat. Commun.* **10**, 5729 (2019).
25. Goodman, A. L., Wu, M. & Gordon, J. I. Identifying microbial fitness determinants by insertion sequencing using genome-wide transposon mutant libraries. *Nat. Protoc.* **6**, 1969–1980 (2011).
26. Barquist, L. et al. A comparison of dense transposon insertion libraries in the *Salmonella* serovars Typhi and Typhimurium. *Nucleic Acids Res.* **41**, 4549–4564 (2013).
27. Christen, B. et al. The essential genome of a bacterium. *Mol. Syst. Biol.* **7**, 528 (2011).
28. Aird, D. et al. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol.* **12**, R18 (2011).
29. Mitra, A., Skrzypczak, M., Ginalski, K. & Rowicka, M. Strategies for achieving high sequencing accuracy for low diversity samples and avoiding sample bleeding using illumina platform. *PLoS One* **10**, e0120520 (2015).
30. Schlechter, R. O. et al. Chromatic bacteria—a broad host-range plasmid and chromosomal insertion toolbox for fluorescent protein expression in bacteria. *Front. Microbiol.* **9**, 3052 (2018).
31. Truong, L. & Ferre-D'Amare, A. R. From fluorescent proteins to fluorogenic RNAs: tools for imaging cellular macromolecules. *Protein Sci.* **28**, 1374–1386 (2019).
32. Lambert, T. J. FPbase: a community-editable fluorescent protein database. *Nat. Methods* **16**, 277–278 (2019).
33. Day, R. N. & Schaufele, F. Fluorescent protein tools for studying protein dynamics in living cells: a review. *J. Biomed. Opt.* **13**, 031202 (2008).
34. Roederer, M. Spectral compensation for flow cytometry: visualization artifacts, limitations, and caveats. *Cytometry* **45**, 194–205 (2001).
35. Williamson, N. R. et al. Biosynthesis of the red antibiotic, prodigiosin, in *Serratia*: identification of a novel 2-methyl-3-n-amyl-pyrrole (MAP) assembly pathway, definition of the terminal condensing enzyme, and implications for undecylprodigiosin biosynthesis in *Streptomyces*. *Mol. Microbiol.* **56**, 971–989 (2005).
36. Andreyeva, I. N. & Ogorodnikova, T. I. Pigmentation of *Serratia marcescens* and spectral properties of prodigiosin. *Microbiology* **84**, 28–33 (2015).
37. del Solar, G., Giraldo, R., Ruiz-Echevarria, M. J., Espinosa, M. & Diaz-Orejas, R. Replication and control of circular bacterial plasmids. *Microbiol. Mol. Biol. Rev.* **62**, 434–464 (1998).
38. Melamed, S. et al. Global mapping of small RNA-target interactions in bacteria. *Mol. Cell* **63**, 884–897 (2016).
39. Paige, J. S., Wu, K. Y. & Jaffrey, S. R. RNA mimics of green fluorescent protein. *Science* **333**, 642–646 (2011).
40. Ouellet, J. RNA fluorescence with light-up aptamers. *Front. Chem.* **4**, 29 (2016).
41. Barnard, A., Wolfe, A. & Busby, S. Regulation at complex bacterial promoters: how bacteria use different promoter organizations to produce different regulatory outcomes. *Curr. Opin. Microbiol.* **7**, 102–108 (2004).
42. Thoma, S. & Schobert, M. An improved *Escherichia coli* donor strain for diparental mating. *FEMS Microbiol. Lett.* **294**, 127–132 (2009).
43. Jackson, S. A., Fellows, B. J. & Fineran, P. C. Complete genome sequences of the *Escherichia coli* donor strains ST18 and MFD*pir*. *Microbiol. Resour. Announc.* **9**, e01014-20 (2020).
44. Patterson, A. G., Chang, J. T., Taylor, C. & Fineran, P. C. Regulation of the Type I-F CRISPR-Cas system by CRP-cAMP and GalM controls spacer acquisition and interference. *Nucleic Acids Res.* **43**, 6038–6048 (2015).
45. Mesarich, C. H. et al. Transposon insertion libraries for the characterization of mutants from the kiwifruit pathogen *Pseudomonas syringae* pv. actinidiae. *PLoS One* **12**, e0172790 (2017).
46. Monson, R. et al. A plasmid-transposon hybrid mutagenesis system effective in a broad range of Enterobacteria. *Front. Microbiol.* **6**, 1442 (2015).
47. Karsi, A. & Lawrence, M. L. Broad host range fluorescence and bioluminescence expression vectors for Gram-negative bacteria. *Plasmid* **57**, 286–295 (2007).
48. Stuurman, N. et al. Use of green fluorescent protein color variants expressed on stable broad-host-range vectors to visualize rhizobia interacting with plants. *Mol. Plant Microbe Interact.* **13**, 1163–1169 (2000).
49. Garcia-Cayuela, T. et al. Fluorescent protein vectors for promoter analysis in lactic acid bacteria and *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **96**, 171–181 (2012).
50. Charpentier, E. et al. Novel cassette-based shuttle vector system for gram-positive bacteria. *Appl. Environ. Microbiol.* **70**, 6076–6085 (2004).
51. Aymanns, S., Mauerer, S., van Zandbergen, G., Wolz, C. & Spellerberg, B. High-level fluorescence labeling of gram-positive pathogens. *PLoS One* **6**, e19822 (2011).
52. Barbier, M. & Damron, F. H. Rainbow vectors for broad-range bacterial fluorescence labeling. *PLoS One* **11**, e0146827 (2016).
53. Birkholz, N., Fagerlund, R. D., Smith, L. M., Jackson, S. A. & Fineran, P. C. The autoregulator Aca2 mediates anti-CRISPR repression. *Nucleic Acids Res.* **47**, 9658–9665 (2019).
54. Genove, G., Glick, B. S. & Barth, A. L. Brighter reporter genes from multimerized fluorescent proteins. *Biotechniques* **39**, 814 (2005). 816, 818 passim.
55. 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).

56. Baba, T. et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006.0008 (2006).
57. Kitagawa, M. et al. Complete set of ORF clones of *Escherichia coli* ASKA library (a complete set of *E. coli* K-12 ORF archive): unique resources for biological research. *DNA Res.* **12**, 291–299 (2005).
58. Peters, J. M. et al. Enabling genetic analysis of diverse bacteria with Mobile-CRISPRi. *Nat. Microbiol.* **4**, 244–250 (2019).
59. Wickham, H. et al. Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
60. Andrews, S. *FastQC: A Quality Control Tool for High Throughput Sequence Data* (Babraham Bioinformatics, Babraham Institute, 2010).
61. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
62. Quinlan, A. R. BEDTools: the Swiss-Army tool for genome feature analysis. *Curr. Protoc. Bioinform.* **47**, 11.12.11–11.12.34 (2014).
63. The R Foundation. R: A Language and Environment for Statistical Computing. https://www.r-project.org/ (2020).
64. Jackson, S. A. JacksonLab/SorTn-seq: 1.0.0 (Version v). Zenodo https://doi.org/10.5281/zenodo.4554398 (2021).
65. RStudio. RStudio: Integrated Development for R. http://www.rstudio.com/ (2020).
66. Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**, 863–864 (2011).
67. Rutherford, K. et al. Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944–945 (2000).
68. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
69. Ponstingl, H. & Ning, Z. SMALT-a new mapper for DNA sequencing reads. *F1000Posters* **1**, 313 (poster) (2010).
70. Robinson, M. D. & Smyth, G. K. Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics* **23**, 2881–2887 (2007).
71. Robinson, M. D. & Smyth, G. K. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics* **9**, 321–332 (2008).

Acknowledgements

This work was supported by the Marsden Fund from the Royal Society of New Zealand and the School of Biomedical Sciences Bequest Fund from the University of Otago. L.M.S. was supported by a University of Otago Doctoral Scholarship and Postgraduate Publishing Bursary. We thank M. Wilson for help with cell sorting and A. Jeffs of the Otago Genomics Facility for assistance with sequencing. We thank J. Ussher and R. Hannaway for help with flow cytometry. We thank H. Hampton, D. Mayo-Muñoz, N. Birkholz and members of the Fineran laboratory for helpful discussions.

Author contributions

L.M.S. designed all experiments with input from S.A.J. and P.C.F. L.M.S. performed all experiments (except where a specialized facility was used), analyzed all data and prepared all figures. S.A.J. wrote data processing scripts, and P.P.G. provided input into bioinformatic analyses. P.C.F. conceived the project. S.A.J., P.P.G. and P.C.F. supervised the project. L.M.S. and P.C.F. wrote the manuscript. All authors edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41596-021-00582-6.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41596-021-00582-6.

Correspondence and requests for materials should be addressed to P.C.F.

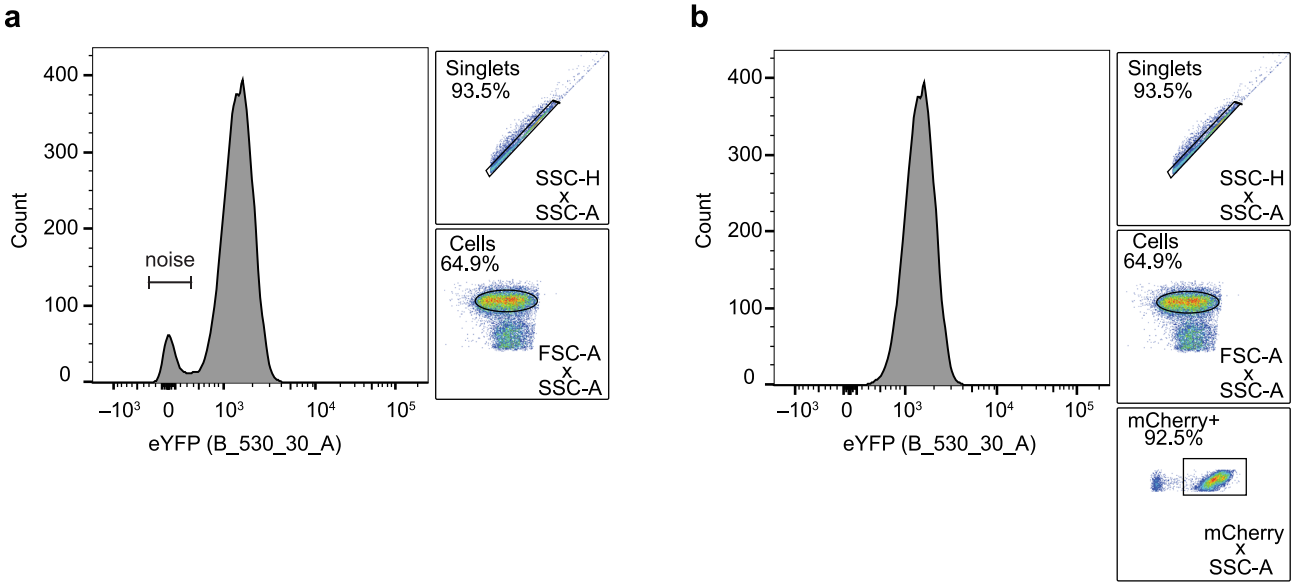
Peer review information *Nature Protocols* thanks the anonymous reviewers for their contribution to the peer review of this work.

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

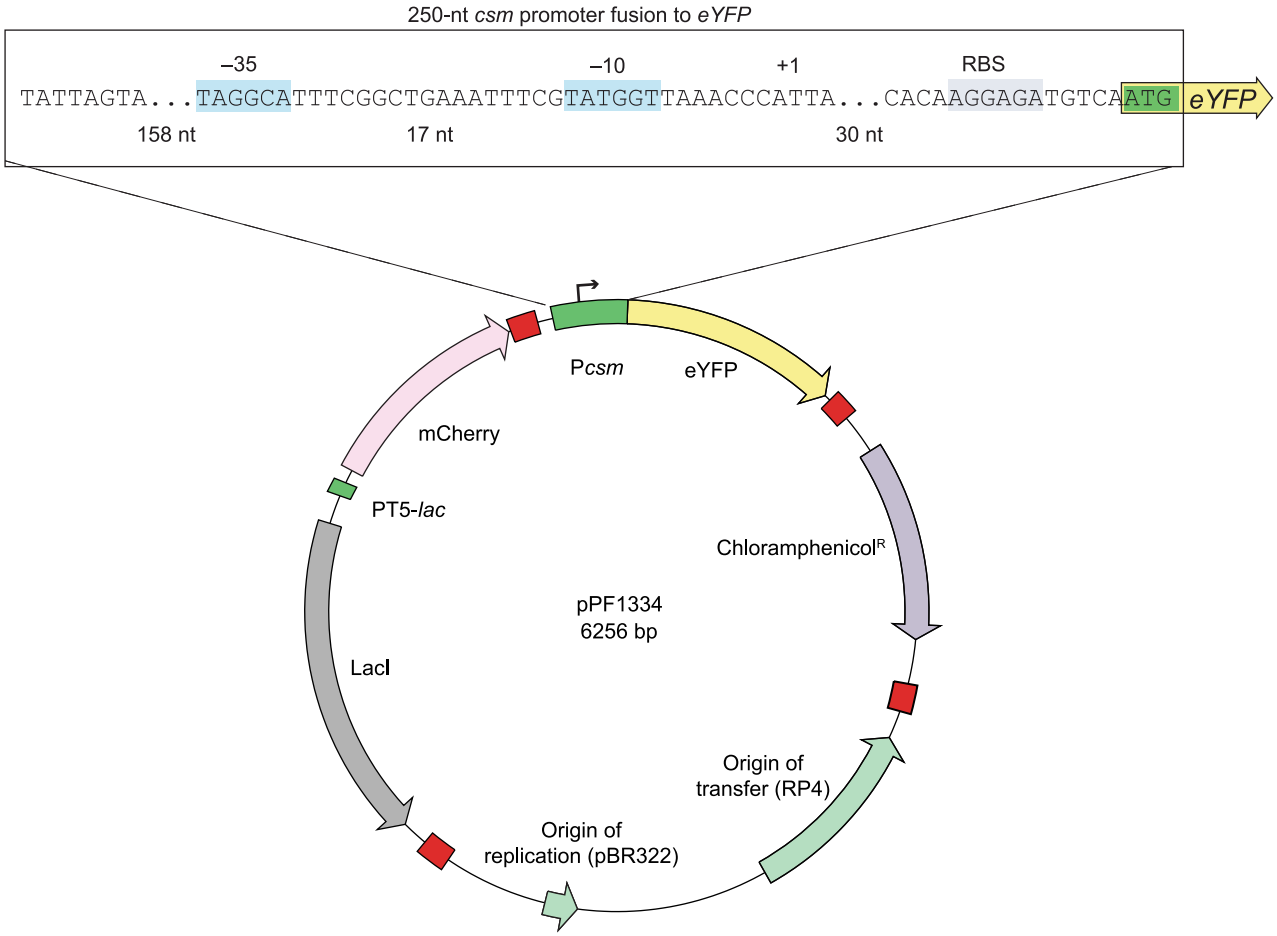
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 10 December 2020; Accepted: 3 June 2021;
Published online: 4 August 2021

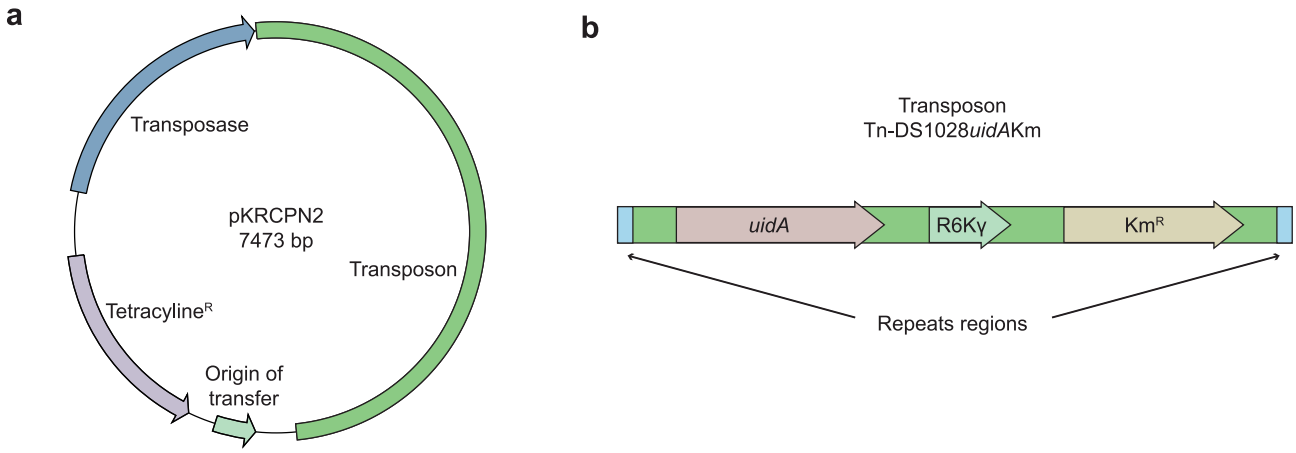
Related links
Key reference using this protocol
Smith, L. et al. *Nat. Microbiol.* **6**, 162–172 (2021): https://doi.org/10.1038/s41564-020-00822-7



Extended Data Fig. 1 | Gating on a secondary reporter eliminates noise from the primary reporter fluorescence distribution. The primary reporter (eYFP) fluorescence distribution is shown for without (**a**) or with (**b**) gating on a secondary reporter (mCherry). In **a** and **b**, events are first gated on SSC and FSC parameters area (A) and height (H) to isolate individual bacteria ('singlets' and 'cells'). **a**, A characteristic secondary peak is observed centered around zero. Many of the events comprising this secondary peak exhibit negative fluorescence levels and therefore cannot be sorted. This non-fluorescent population is probably comprised of dead/dormant cells, cellular debris, bubbles or electronic noise generated by the instrument. **b**, The addition of a gate around mCherry-positive (mCherry+) events results in the removal of the secondary peak from the eYFP fluorescence distribution. This improved distribution allows for more accurate placement of gates during cell sorting.



Extended Data Fig. 2 | Promoter region and reporter plasmid used during SorTn-seq. Type III-A CRISPR-Cas expression was measured from an eYFP fusion to 250 nt upstream of the *csm* operon. Key features in the promoter are indicated (−35 and −10) as are the transcription start site (+1), native RBS and the start codon (ATG, green). An IPTG-inducible 2nd fluorophore (mCherry) is under the control of the T5-*lac* promoter (PT5-*lac*). The pBR322 origin of replication facilitates ~15–20 copies of the reporter per cell.



Extended Data Fig. 3 | Transposon delivery vector and transposon organization. a, Organization of the pKRCPN2 transposon delivery vector. **b**, Schematic of the transposon Tn-DS1028*uidA*Km, which contains a transcriptional *uidA* reporter, origin of replication (R6Kγ) and kanamycin resistance gene (Km^R).