

Reflecting on a Decade of Transposon-Insertion Sequencing

Amy K. Cain^{1*}, Lars Barquist^{2,3}, Andrew L. Goodman⁴, Ian T. Paulsen¹, Julian Parkhill⁵ and Tim van Opijnen^{6*}

Authors Affiliations:

¹ARC Centre of Excellence in Synthetic Biology, Department of Molecular Sciences, Macquarie University, Sydney, NSW, Australia

² Helmholtz Institute for RNA-based Infection Research, Helmholtz Centre for Infection Research, Würzburg, BY, Germany

³ Faculty of Medicine, University of Würzburg, Würzburg, BY, Germany

⁴ Department of Microbial Pathogenesis and Microbial Sciences Institute, Yale University School of Medicine, New Haven, CT, USA

⁵Department of Veterinary Medicine, University of Cambridge, Cambridge, Cambs, UK

⁶ Department of Biology, Boston College, Boston, MA, USA

*corresponding authors: amy.cain@mq.edu.au or vanopijn@bc.edu

19 **Abstract**

20 It has been 10 years since the birth of modern transposon-insertion sequencing (TIS)
21 methods, which combine genome-wide transposon mutagenesis with high-throughput
22 sequencing to estimate the fitness contribution or essentiality of each genetic component
23 simultaneously in a bacterial genome. Four TIS variations were published in 2009:
24 transposon sequencing (Tn-Seq), transposon-directed insertion site sequencing (TraDIS),
25 insertion sequencing (INSeq) and high-throughput insertion tracking by deep sequencing
26 (HITS). TIS has since become an important tool for molecular microbiologists, being one of
27 the few genome-wide techniques that directly links phenotype to genotype and ultimately can
28 assign gene function. In this review, we discuss the recent applications of TIS to answer
29 overarching biological questions. We explore emerging and multi-disciplinary methods that
30 build on TIS, with an eye towards future applications.

32 **Introduction**

33 Transposon-insertion sequencing (TIS) methods combine large-scale transposon mutagenesis
34 with next-generation sequencing to estimate the essentiality and/or fitness contribution of
35 each genetic feature in a bacterial genome simultaneously. A strength of TIS is that
36 experiments are performed with pooled transposon libraries, which allows direct linkage of
37 phenotype to genotype in a high-throughput manner. Ultimately TIS aims to elucidate the
38 function of each genomic feature and is therefore a critical tool to help interpret the mounting
39 levels of genome sequencing data being generated. TIS methods can be sensitive enough to
40 detect even minor changes in mutant fitness but also, with sufficient density, precise enough
41 to be able to assay not only genes, but intergenic regions, promoter regions, and essential
42 protein domains within coding regions. Four variations on the TIS method were published
43 simultaneously in 2009: transposon sequencing (Tn-Seq)¹, transposon-directed insertion site
44 sequencing (TraDIS)², insertion sequencing (INSeq)³ and high-throughput insertion tracking
45 by deep sequencing (HITS)⁴. Since then, TIS has become a valuable tool in our molecular
46 biology toolkit, whose full utility is still being explored.

47
48 The basic TIS workflow is summarised in Figure 1. Briefly, it begins with construction of a
49 saturated mutant library (Fig 1A) by introducing a randomly-inserting transposon (Tn),
50 commonly a Tn5 or *mariner* Tn, into a strain of interest often using transformation or
51 conjugation. The goal is to create a population of bacteria where each cell carries a single Tn

52 insertion in the genome, and each genetic component is disrupted multiple times at different
53 sites. By directly sequencing the transposon flanking regions of the initial library, potential
54 essential features can be identified as those that do not tolerate insertions. Alternatively, the
55 library can be subjected to a selective condition, for instance antibiotic stress (Fig 1B), in
56 order to query features involved in survival and growth within that environment. Such
57 conditionally important components are defined by insertions whose frequency significantly
58 changes in the population during the selection, determined by sequencing before and after
59 selection. Genomic features that have Tn insertions with a decrease in frequency over
60 experimental selection are assumed to be important for fitness in the test conditions; such
61 features could include antibiotic resistance genes during antibiotic selection, or virulence
62 factors in an infection model. Features where insertions show an increase in frequency are
63 assumed to have an advantageous effect in the test conditions; including negative regulators
64 or metabolically-costly systems not necessary in those conditions.

65

66 There are 4 major TIS versions that differ in various steps of their sequencing procedures (see
67 REF⁵ for more detail on these variations). For example, the way DNA is fragmented for
68 library preparation varies: two methods (Tn-Seq and INSeq) use the type II restriction
69 enzyme MmeI to yield uniform-length shorter reads, which can remove PCR amplification
70 bias, while the other two (TraDIS and HITS) use random-sized shearing via sonication, that
71 can have the advantage of improved Tn mapping owing to longer reads. Similarly, Tn-Seq
72 and INSeq exclusively use the mariner Tn, which inserts into TA (thymine-adenine
73 dinucleotide) but otherwise does not have a sequence preference, and the others have the
74 flexibility that they can employ any Tn, but commonly use Tn5 as it is commercially
75 available and does not have a insertion site bias. After fragmentation, various adaptors are
76 added, and transposon–genome junctions are amplified and sequenced with a sequencing
77 primer either facing out of the Tn or adapter. Finally, mapping of the adjacent genomic DNA
78 allows the exact position of each transposon in the bacterial genome to be determined using
79 appropriate bioinformatic tools (see the “Developments in TIS data analysis” section below).

80

81 Since the last comprehensive reviews on TIS^{5,6} in 2013, a range of exciting and multi-
82 disciplinary methods that build on TIS have emerged to answer increasingly complex
83 biological questions. These include scaling TIS analysis to hundreds of different conditions
84 using high-throughput phenotyping, the use of machine learning to predict bacterial survival
85 outcomes, and combining transposon-insertion sequencing with cutting-edge techniques from

single-cell analysis (droplet Tn-Seq) and fluorescence sorting (TraDISort). Analysis tools have also evolved to cope with this increase in complexity of TIS studies. Lastly, a broad range of *in vitro* and *in vivo* applications of TIS have been implemented in pathogenic, commensal and environmental bacteria in the last decade. In this review, we discuss these exciting developments and applications of TIS and present our vision for TIS into the future. Detailed information on the experimental design of TIS experiments, including choice of Tn and statistical impacts of experimental parameters, comparisons of TIS method variations, limitations of standard TIS and details on applications prior to 2013, can be found in previous reviews⁷⁻⁹.

Advances and extensions of TIS methods

Over the last decade, TIS methods have been developed to incorporate other technologies and techniques to answer complex biological questions in creative ways. These include physical separation and sorting of individual mutant cells, using inducible promoters to study essential genes and scaling of current techniques to simultaneously screen multiple environments and different species enabling pan-organism analysis (see Fig 2).

Beyond growth-based selection approaches. A major recent advance of TIS is based on the ability to separate mutants by their physical characteristics, rather than selecting solely on growth. The simplest forms of this have adapted classical microbiological assays to the massive multiplexing enabled by TIS (Fig 2A). For example, motility genes can be assayed by “racing” mutant libraries across agar plates and comparing mutants in the inner (less motile) to the outer (more motile) populations. This approach has been applied to *Escherichia coli* ST131¹⁰ and *Pseudomonas aeruginosa* PA14¹¹, leading to the identification of known motility genes, like those encoding common bacterial motors (flagella, fimbria and pili), in addition to new candidates. Similarly, Density-TraDISort¹² combines TraDIS and density gradient centrifugation to separate mutants based on their density (Fig 2B) and identify genes involved in bacterial capsule production, a major virulence factor for many pathogens. In this study, 78 genes underlying capsule production were identified across two clinically relevant *Klebsiella pneumoniae* strains¹².

The application of cell sorting to TIS has led to the development of techniques that progress from bulk separation to separation at the level of single cells. One such application is

TraDISort, which combines fluorescence-activated cell sorting (FACS) and TraDIS¹³ and sorts single cells based on fluorescence. TraDISort has used the cytosolic concentration of ethidium bromide (EthBr), a fluorescent DNA intercalating agent, as a marker for altered efflux activity (Fig 2C). For instance, mutants with insertions in efflux pump genes, such as *amvA* had reduced ability to remove EthBr from the cell, resulting in an overall higher level of fluorescence. In contrast, mutants such as the *amvA*-repressor (*amvR*), had increased efflux and lower fluorescence. A similar approach used a fluorescent reporter to separate heterogeneous populations of *Mycobacterium tuberculosis*¹⁴, which uncovered *lamA*, a gene of previously unknown function that reduced overall heterogeneity in the population by decreasing asymmetric polar growth. Similarly, FAST-INSeq was developed to identify regulators of Typhoid toxin production, using FACS of *Salmonella* Typhi-infected macrophages with a fluorescent reporter for toxin expression¹⁵. Lastly, Tn-FACSeq was used to identify genes from *Bdellovibrio bacteriovorus*, a bacterial predator, that are important for attachment to *Vibrio cholerae*¹⁶. These types of fluorescence-based techniques could be extended further, for instance to examine bacterial responses to other fluorescent (or fluorescently-tagged) compounds, other fluorescent reporter constructs or simply using FACS to differentiate mutants with altered cell size.

Population-independent mutant assays

In traditional TIS approaches mutant fitness is measured within the context of the entire mutant population. However, the true fitness of a mutant can be obscured when grown in the presence of other mutants. For instance, TIS cannot report on the effect of secreted products or other ‘common goods’ which act beyond the cell containing the mutation, or similarly mutants that suffer from density dependence. Recently droplet Tn-Seq (dTn-Seq) has been developed to address these issues, based on advances in single-cell analysis¹⁷. dTn-Seq sorts single mutants by combining microfluidics with TIS, encapsulating individual transposon mutants in growth medium-in-oil droplets, enabling isolated growth of mutants free from the influence of the population (Fig. 2D). dTn-Seq experiments showed that in *Streptococcus pneumoniae* 1-3% of mutants have altered fitness when grown in isolation; some mutants may grow faster or slower in isolation, compared to their fitness measured in a traditional TIS screen. To highlight its versatility, dTn-Seq has been applied to investigate hyper-competence, processing of host glycoproteins, defence against host immune factors and microcolony formation¹⁷. Moreover, dTn-Seq is compatible with microscopy and FACS-based screening, and by reloading droplets into a microfluidics device, multi-layer

encapsulations can be achieved. Such droplets consist of multiple layers containing different mutants, other bacterial species or even host cells, between which communication signals can freely diffuse, thereby enabling investigations of inter-bacterial and bacteria-host cell interactions (Fig2D)¹⁷. Although these droplets cannot easily be applied to *in vivo* animal models, any interesting phenotypes that arise from dTn-Seq screens, including host-microbe interaction mediators, can be directly confirmed in cell culture assays and/or *in vivo* in animal models using targeted mutants.

Assaying function of essential genes and gain-of-function screens. One limitation of standard TIS is that only non-essential genes can be assayed, as essential genes, by definition, do not tolerate insertions. A handful of studies have overcome this by using gain-of-function screens that utilize libraries of transposons with outward-facing promoters to facilitate gene overexpression and repression (Fig 3A). Monitoring the change in frequency of transposons that induce the expression of downstream genes, including essential genes, during selection can identify phenotypes that may not be evident from gene disruption (Fig 3B). This idea is not new; for example the TnAraOut method, developed in 2000¹⁸, used transposons containing the arabinose inducible promoter P_{BAD} to screen for essential antibiotic targets. Various approaches, where an outward facing promoters are engineered into a transposon system, have been developed to assay essential genes, for example in *Caulobacter crescentus*¹⁹ and *Staphylococcus aureus*^{20,21}. Recently, this approach was combined with traditional TIS, resulting in the TraDIS-Xpress package²² (previously known as TraDIS+²³). TraDIS-Xpress uses an inducible P_{BAD} promoter facing out of a Tn5 in *E. coli*, in addition to detailed transposon inactivation data, to query all genes. It was successfully applied to identify both essential and non-essential genes affecting tolerance to varying concentrations of the biocide triclosan, finding differential responses to bactericidal and bacteriostatic concentrations. A high-throughput method for gain-of-function assays was recently developed, dual-barcoded shotgun expression library sequencing (Dub-Seq), where barcoded overexpression libraries of *E. coli* were mapped, barcoded and used to assign gene function in 52 experimental conditions based on mutant fitness changes due to increased gene dosage²⁴. To control expression in gain-of-function screens, some studies^{19,22} employed inducible promoters on a single transposon, which can have the advantage (over using multiple transposons) of allowing high library density and reducing insertion bias. Other studies use constitutive promoters with different strengths on either barcoded or different types of

transposons, which has the strategic advantage that different gene dosages can be assayed in the same culture^{20,21}.

Scaling up TIS using high-throughput phenotyping. While it is possible to apply the original TIS protocols at scale, the multi-step library preparations involved can become increasingly costly when dealing with hundreds of samples. One solution to this problem, Random Barcode Transposon-site Sequencing (RB-Tn-Seq)²⁵ introduces a random DNA barcode into each transposon. An initial conventional TIS sequencing approach is used to determine the insertion site associated with each barcode, and then a single-step PCR barcode amplicon can be directly sequenced in future experiments to track changes in mutant frequencies²⁶, significantly speeding up screening. For instance, one recent up-scaled study applied RB-Tn-Seq to 32 different bacterial strains across 129 conditions, identifying a large variety of leads for gene function²⁷. A second problem in scaling TIS to large collections of bacteria is that optimized transposon delivery vectors often do not exist for non-model organisms. The ‘magic pool’ approach accelerates the optimization process using pools of transposon vectors, each of which has a different combination of upstream sequences (promoters and ribosome binding sites) and antibiotic resistance markers as well as a random DNA barcode sequence, which allows quick measurement of vector efficiency during mutagenesis²⁸.

Developments in TIS data analysis

A typical analysis protocol of TIS data can be summarised as follows: after splitting sequencing reads based on their multiplexing barcode, any transposon or adapter sequences are removed, reads are mapped to an annotated reference genome, and the unique position of the transposon and relative insertion coverage (i.e. number of reads) are recorded, processed and presented to calculate the effect of each Tn on fitness (e.g. growth or survival). Over the past 10 years, a number of bioinformatics pipelines and protocols have been developed for this purpose (Table 1). These TIS tools include web-based applications^{29 30}, standalone graphical applications³¹, and command-line toolkits^{32,33 34}. All of these tools implement variations on both gene essentiality and conditional fitness analyses, although they differ in the details of preprocessing and read alignment, normalization techniques, and statistical models or tests used. Often choices made in the experimental protocol can impact the appropriateness of a particular analysis procedure as much as any theoretical issue; for

instance, many hidden Markov model and sliding window approaches to defining essential regions in the absence of annotation are only applicable to mariner transposon studies, as the assumption of a uniform insertion probability at TA sites simplifies the underlying statistical model. Many of these core issues have been addressed in a 2016 review on design and analysis of TIS experiments⁷, but recent TIS developments have pushed analysis methods in new directions.

One major development of TIS analysis methods is in dealing with infection dynamics and particularly the effects of bottlenecks, which are transient reductions in population size during the course of the experiment (see below). While these have been dealt with in analyses using normalization based on changes in neutral loci³⁵ or subsampling³², two interesting new approaches to this problem have recently been proposed. In the first, principal component analysis (PCA) is performed on log fold changes (logFCs) in mutant abundance across replicate infection experiments³⁶. Examination of the principal components recovered can then identify linear combinations of the changes across replicates that separate genes consistently across experiments, providing a score for association of any particular gene with survival in infection and eliminating the contribution of spurious stochastic changes. A second approach has adopted the zero-inflated negative binomial (ZINB) distribution to model transposon insertion counts³⁷. The ZINB distribution is a mixture of a logit distribution, which captures the probability of detection of a data point, and the negative binomial, which captures the overdispersion generally observed in sequencing data. This distribution has attracted recent attention in the analysis of single-cell RNA sequencing (RNA-seq), as it provides a natural mechanism for capturing technical dropout of transcripts³⁸. Similarly, by fitting the logit component of the ZINB genome-wide, this approach can correct for differences in library saturation between conditions arising either in library creation or due to bottlenecking.

A second development is the move from simple condition:control comparisons to the simultaneous investigation of large suites of conditions. For instance, the PCA and ZINB methods highlighted above demonstrated their effectiveness using combinations of existing datasets, identifying commonalities in genes required by different *Vibrio* strains in infection³⁹ or response to a panel of antibiotics in *M. tuberculosis*³⁷, respectively. A striking example of such a data analysis pipeline, named AlbaTraDIS⁴⁰ was developed for the TraDIS-Xpress study examining triclosan tolerance (see REF²² and the TraDIS-Xpress discussions above)

and uses sliding window analyses integrating all available information to predict all genes and promoter regions involved during selection.

As TIS experiments become more complex and are combined with other datatypes, tools for visualization and data delivery are becoming increasingly important. For instance, a recent study integrating expression and fitness data from the HIV-associated *Salmonella enterica* serovar Typhimurium strain D23580⁴¹ included a Dalliance-based browser⁴² allowing readers to directly interrogate the data themselves, providing a valuable community resource. Platforms for easily providing this kind of interactive interface are beginning to emerge, such as ShinyOmics⁴³, a web-based application for rapid collaborative exploration of omics-data including TIS, RNA-seq and proteomics, enabling comparisons between datasets, PCA, and simple network analysis. As datasets accumulate and automation increases throughput, such integrative analysis approaches will become increasingly important.

Key biological applications of TIS

Since its development, TIS has been used in a range of *in vitro* studies as well as *in vivo* infection models. Here we summarise how the development of TIS has enabled the investigation of key biological questions with a focus on studies with implications for human health.

Identifying genes and networks involved in antibiotic resistance. The emergence of antibiotic resistance is a major global health problem, exacerbated by a lack of development of new antibiotics. TIS is well equipped to infer the relative impact that disrupting each genomic feature has on antibiotic sensitivity (Fig 4) and can contribute to developing a better understanding of how resistance emerges, as well as guide the development of new strategies to target (resistant) bacteria. Traditional TIS experiments performed by culturing transposon libraries with inhibitory but sub-lethal concentrations of antibiotics for several generations have been used to define a comprehensive non-essential gene complement involved in intrinsic resistance for many clinically important pathogens, including the notorious ESKAPE species (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* species)^{20,44–47}.

These studies have shown that while antibiotics may have specific targets (e.g. in cell wall synthesis, DNA replication, or protein synthesis), the bacterial response to antibiotics is actually distributed across the genome. For example, fluoroquinolones target the topoisomerase IV and DNA gyrase enzymes, which are essential enzymes involved in DNA replication. While most TIS experiments cannot assay these targets directly due to their inherent essentiality, TIS profiles generated under fluoroquinolone exposure implicate other genes involved in DNA replication and repair such as *recN* and *xseA*^{35,48}. While these genes are not direct fluoroquinolone targets, they contribute to intrinsic resistance as part of a secondary effect; for example, fluoroquinolones trigger DNA damage, which activates DNA repair. In general, antibiotic TIS profiles for each antibiotic tested, and each organism screened show a role for genes beyond those related to the primary target, indicating the importance of genes with diverse functions including amino acid and carbohydrate metabolism, energy generation, transport and regulation^{49 50 45 51 46 44 52 53 54}. This underlines that while we have a limited view of how an antibiotic inhibits a bacterial cell, TIS can be used to uncover this complex, multifactorial process. As a result, TIS profiles have been demonstrated to be effective in determining the mechanism of action of novel antibiotics^{55,56}. Creating profiles for multiple similar conditions can help direct attention to genes with unknown function that are important under all of these conditions and thereby help identify leads that may help uncover their function (Figure 4A.2 and see section 3.7). Moreover, TIS profiles can also uncover opportunities to sensitize a bacterium to a drug, enabling the design of secondary or helper drugs^{57 44}.

Investigating virulence genes, host-adaptation, and vaccine development. Interrogating the genomic requirements for pathogens to cause disease has been a major motivation in the development of large-scale reverse genetics approaches (Fig 5). For instance, the development of early transposon screens in *Salmonella* Typhimurium^{58 59} provided key evidence for the discovery of major virulence factors. TIS has made it much easier and faster to screen for virulence factors in bacterial pathogens. These have included *in vitro* assays, such as capsule production^{60 12}, growth in serum⁶¹ and colicin resistance and sensitivity in *E. coli*⁶², as well as sporulation in *Clostridium difficile*⁶³. An early *in vivo* screen used retrospective TIS on existing samples to examine *S. Typhimurium* genes involved in infection of 3 farm animals (chickens, cows, and pigs^{64 65}) plus mice, and found multiple conserved virulence genes. Other examples include: an analysis of the virulence genes of *Legionella pneumophila* using both cell culture and murine models⁶⁶; survival of

Streptococcus pyogenes in human saliva⁶⁷; survival of *A. baumannii* in a bloodstream infection murine model⁶⁸; survival of *Streptococcus equi* in horse blood or hydrogen peroxide⁶⁹; demonstrating that oxidative stress resistance enhances *V. cholerae* host adaptation in a murine model⁷⁰; survival of *Burkholderia cenocepacia* in a *Caenorhabditis elegans* host⁷¹; *Streptococcus mutans* infection in an oral rodent model⁷²; and building a targeted sub-library of type-IV secreted proteins in the intracellular pathogen *Coxiella burnetii* using INSeq, which was subsequently screened for vacuole formation in human HeLa cells⁷³.

While TIS has streamlined the process of generating and monitoring mutant libraries, challenges remain in applying the technique to infection models. A primary concern is the effect of bottlenecks, which can be quantified experimentally by measuring loss of neutral markers³⁵, for instance using the wild-type isogenic tagged strains (WITS) method⁷⁴. Bottleneck effects should be considered computationally (see above) and can, at least partially, be avoided by careful consideration of the infection model. The size and temporal structure of a bottleneck is often specific to the particular infection model, and can be influenced by a range of factors including physical barriers, nutrient availability, and competition with native microbiota⁷⁵. While bottleneck effects can be partially compensated for during analysis if mild, major bottlenecks can irreversibly bias an experiment that does not account for them. In these cases, the surviving mutants represent the subset of bacteria that happened to pass some barrier to infection, rather than being representative of all mutants that could, leading to skewed representation and a lack of reproducibility between experimental replicates. The wide variety of infection models developed to study *Salmonella* Typhimurium provide an example of how model choice can affect the design of a transposon screen (Fig 5). For the most realistic infection models based on infection through the gut epithelium, bottlenecks can be severe^{76,77}, limiting transposon analysis to small pools of tens to hundreds of mutants^{64 65}. Intraperitoneal inoculation can bypass this major bottleneck in a mouse model, which has allowed for screening of ~10,000 mutants in a single animal⁷⁸; however, the results are uninformative with regards to gastrointestinal disease. Finally, in the case of cell culture models, such as the macrophage model which captures a key challenge to the development of systemic salmonellosis⁵⁸, the only constraint on library complexity is the number of cells available for infection, allowing for efficient screening of very large mutant populations (~10⁶)⁴¹. Similar trade-offs between realism and library complexity are likely to

exist for many infection models, particularly those that involve bacterial penetration of barrier defences.

Experimental challenges notwithstanding, TIS applied to animal models has also proved useful in identifying and understanding vaccine targets. For example, screening of a *S. pneumoniae* TIS library in a ferret transmission model, describing the fitness landscape of genes during mammalian transmission, yielded valuable and translatable data. Targeted deletion confirmed that key TIS hits (putative C3-degrading protease CppA, iron transporter PiaA, and competence regulatory histidine kinase ComD) significantly decreased transmissibility. Importantly, maternal vaccination with recombinant PiaA and CppA alone or in combination blocked transmission from mother to offspring and was more effective than capsule-based vaccines⁷⁹. In a second example, a murine sickle cell disease (SCD) model coupled with TIS identified a set of pneumococcal virulence genes specific to hosts with SCD. Not only did these factors point to aspects of SCD pathophysiology, but also showed that the protective capacity of antigens can be different in healthy versus the SCD population, highlighting the importance of understanding bacterial pathogenesis in the context of common comorbidities⁸⁰.

Assaying functional components of mobile genetic elements. Mobile genetic elements (MGEs), including plasmids, transposable elements and bacteriophages are important players in inter- and intra-species gene transfer and are heavily implicated in the spread of antibiotic resistance and virulence determinants. Plasmids are notoriously difficult to study with screening approaches, due to their independent replication systems and capacity to regulate copy number. TIS has been used to identify genes involved in maintenance of IncA/C plasmids in *E. coli*, and the results were then developed into an IncA/C plasmid typing scheme⁸¹. Additionally, TIS has also been used to demonstrate the involvement of a type 4 secretion system in conjugation of an IncP plasmid in *Edwardsiella piscicida*⁸².

Bacteriophages ('phages') are important MGEs that have been used for bacterial typing for decades, and can greatly influence bacterial pathogenesis through the transduction of pathogenicity islands. Additionally, phage therapy to treat resistant bacterial infections is experiencing a resurgence as an alternative to antibiotics. TIS is able to identify essential host factors that mediate or hinder phage infection. For example, challenge of an *E. coli* O157 TraDIS library with T4 and T7 bacteriophages identified new host genes involved in both

bacteriophage resistance (e.g *sspA*; stringent starvation protein A) and susceptibility (e.g. *sap* operon)⁸³. Similarly, experiments with phages specific to particular capsules has not only allowed the identification of modifiers of phage resistance⁸⁴, but also genes responsible for capsule expression⁶⁰.

Uncovering essential genes and the influence of the pan-genome. One of the first uses of TIS was to define the essential genes for survival for a plethora of bacterial species, including human pathogens such as *Porphyromonas gingivalis*⁸⁵, *Burkholderia cenocepacia*⁸⁶ and *Yersinia pseudotuberculosis*⁸⁷, animal pathogens such as *S. equi*⁸⁸, plant pathogens such as *Pseudomonas syringae*⁸⁹; the model organism *E. coli* K12⁹⁰ or commensal gut bacteria such as *Bifidobacterium breve*⁹¹. The invaluable essential gene datasets gained from TIS studies have been shown to correlate well with existing phenotypic gene essentiality data, such as from the single gene *E. coli* knock out (KEIO) library⁹² and these can not only be interpreted to understand basic functioning of the cell, but can also provide vital information for identifying potential novel drug targets.

Once TIS is implemented in a species it is often straightforward to create libraries in related strains. This flexibility enables functional exploration of a species' pan-genome and how genetic background can affect phenotype. This kind of in-depth investigation can ultimately help to uncover how bacterial species, particularly diverse species that include both pathogens and non-pathogens, can become harmful or antibiotic resistant, for example via horizontal transfer of pathogen-associated genetic material. TIS in 9 strains of *P. aeruginosa* determined the core essential genome in five media, and highlighted that essentiality of some genes depends on genomic context⁹³. The influence of genetic background on phenotype is further illustrated by examples that highlight how genes involved in responding to antibiotic stress can be strain-specific. For example, screening two *S. pneumoniae* isolates for genes that are important for intrinsic resistance to antibiotics from three classes showed that on average only ~50% of the responsive genes are common between strains. Investigating the underlying reasons for this shows that network architecture, including regulatory pathways that direct competence, are wired in a strain-specific manner thereby making responses strain-specific⁵⁷. A recent study probed 5 diverse strains of *S. aureus* for daptomycin resistance mediators and identified several core pathways consistently involved across strains, including the lipoteichoic acid pathway, as well as factors that varied with strain diversity, like cell envelope²⁰. Furthermore, in *Acinetobacter baumannii* a single *gyrA*

resistance allele results in preferential poisoning of topoisomerase IV by ciprofloxacin, leading to large alterations in the fitness landscape of insertion mutants compared to a wild-type *gyrA* background. This altered background triggers the activation of prophage, and quickly leads to the emergence of ciprofloxacin-resistant clones⁹⁴. In *Mycobacterium tuberculosis*, loss-of-function mutations in *katG* can result in isoniazid resistance. However, TIS experiments have shown that several clinical strains have an increased requirement for the gene encoding KatG, compared to the reference strain H37Rv⁹⁵. This variability underscores how genome variation can affect adaptive solutions, and highlights the importance of extending TIS to clinical isolates which may have a very different genetic background to lab strains.

Understanding metabolism, the response to environmental factors and microbe-microbe interactions. Despite decades of accumulating genome sequences in public databanks, many protein coding genes remain unannotated or carry inaccurate annotations, particularly in non-model and difficult-to-culture organisms. Often even the conditions under which a gene contributes to survival are unknown, leading to a serious roadblock in any attempt at molecular characterization. TIS can help us understand how bacterial cells experience the changing environments encountered in nature, by providing comprehensive profiling of mutant phenotypes⁹⁶. Specifically, recent TIS studies have identified fitness determinants during energy-limited growth in *Pseudomonas aeruginosa*⁹⁷, and outlined how a number of bacterial species synthesize amino acids⁹⁸. Others have identified genes in *E. coli* that promote survival during exposure to ionizing radiation⁹⁹, genes in *S. Typhi* that allow adaptation to survival in water¹⁰⁰; and genes involved in desiccation stress in *Salmonella* Typhimurium¹⁰¹. TIS approaches have also uncovered how bacteria interact with plants or deal with soil environments¹⁰². For instance, TIS was used to identify genes required for growth of the soft-rot pathogenic bacterium *Dickeya dadantii* in chicory plants¹⁰³, or genes of *Pantoea stewartii* essential for survival *in planta* to provide insights into how it causes wilt disease in corn¹⁰⁴. As the breadth of examined stresses increases, shared adaptations to diverse stress conditions are emerging.

In a massively up-scaled example of assaying genes for adapting to changing environment in parallel, a recent study demonstrated the utility of this approach by applying RB-Tn-seq to 32 diverse bacteria in ~200 conditions; this work assigned a phenotype to >11,000 uncharacterized proteins, with ~2,000 of these functional annotations demonstrated to be

conserved across organisms²⁷. A similar study investigated the major human gut commensal and obligate anaerobe *Bacteroides thetaiotaomicron* across 492 conditions, identifying genes involved in metabolism and bile tolerance¹⁰⁵. Together, these two studies demonstrate how TIS can be applied broadly across organisms, and deeply within an organism to extract leads for future molecular characterization.

TIS approaches can also directly answer questions with implications for human health; for instance, how does the gut microbiota affect drug metabolism and toxicity? Using the robotics driven TIS mutant array approach pioneered by the INSeq method^{3 106}, where individual mutants are mapped through combinatorial pooling, a comprehensive library of ~1,300 *B. thetaiotaomicron* mutants were selected and incubated with the antiviral drug brivudine. Drug metabolites were then measured by mass spectrometry, and individual *B. thetaiotaomicron* genes required for the production of a hepatotoxic metabolite were identified. Follow-up studies in gnotobiotic mice confirmed the *in vivo* relevance of this toxin production pathway¹⁰⁷, illustrating the power of this approach in understanding drug–microbiota interactions.

In the wild, microbes rarely live planktonically in isolation, but are constantly interacting with other microbes and forming communities, either by chance as in wound co-infections, or as part of a stable ecosystem as in the mammalian gut. TIS provides an opportunity to understand these microbe–microbe interactions, as the genetic response of one bacterium to others can be recorded. Numerous studies have shown that the fitness effects of gene disruption can depend critically on the presence of other community members. For instance, co-infection has been shown to alter the bacterial fitness landscape in wound models, such as with the opportunistic pathogens *Streptococcus gordonii* and *Aggregatibacter actinomycetemcomitans*¹⁰⁸. Furthermore, studies of co-infection with *P. aeruginosa* and *S. aureus* in mouse surgical wounds showed that ~25% of *S. aureus* genes essential during co-infection are no longer needed during single species infection (Fig 6a). Further, single mutants, like of the community dependent essential gene *udk*, encoding a uridine kinase, were confirmed to influence levels of co-infection, but not mono-infection *in vivo*¹⁰⁹. Interaction studies have even been extended to predatory relationships, such as in studies of *Bdellovibrio bacteriovorus* that identified genes required for predation of *V. cholerae* during planktonic and biofilm growth, using Tn-Seq libraries of both the prey¹¹⁰ and predator¹¹¹. Similarly bacterial genes that influence infection with viruses have been examined using TIS

in a number of bacterial species^{60,83,84,112}. The mechanisms of bacterial interactions within communities have also been investigated by screening for effectors of Type VI protein secretion systems (T6SSs). T6SSs are conserved bacterial defence mechanisms, that deliver toxins to neighbouring cells through a contact-dependent mechanism, killing those which lack immunity proteins (Fig 6b). New toxin and immunity genes have been identified through TIS in *V. cholerae*¹¹³ and *P. aeruginosa*¹¹⁴. Together, these studies have provided insight into a wide range of relationships that shape the microbial environment.

Integrating TIS with other genomic approaches to predict complex traits.

While TIS has been most commonly used to make simple associations between environments and genetic components, it can also uncover more complex relationships. No genomic element, gene or pathway exists in isolation; rather they are connected through intricate networks resulting in specific organismal properties and, ideally, an appropriate response when disturbed. One layer of these networks is gene regulation, including the non-coding genome. By combining saturated TIS libraries with expression profiling, functional non-coding RNAs (ncRNAs) can be identified. For example, RNA-seq can be used to map out expression units across the entire genome, and indicate whether non-coding/intergenic regions display significant levels of transcription. In turn, a parallel TIS experiment with insertions in these regions can then be used to associate a phenotype with the disrupted ncRNA. This approach was used in *S. pneumoniae*, which yielded 89 ncRNAs, over half of which had not been identified previously, and several could be associated through TIS *in vivo* data, as being critical for virulence¹¹⁶. A follow-up study employed different RNA-Seq techniques, including TERM-Seq, 5' and 3'-end Seq, to map the full transcriptional landscape of the *S. pneumoniae* virulent type strain TIGR4¹¹⁷. This resulted in identification of a large number of non-coding regulatory regions, which could be associated with a phenotype through integration of TIS data from different environments. Another comprehensive TIS-based regulatory study in *Neisseria meningitidis* identified 288 genes and small ncRNAs needed for colonization of human epithelial and/or endothelial cells¹¹⁸.

Mapping of genetic interactions (GIs), which quantify fitness dependencies between genes, can be used to build genetic interaction networks and infer regulatory relationships, pathway structures, or leads for gene function¹¹⁹. GIs can be identified by creating a TIS library in a query gene deletion background, and screening for genes whose fitness deviates from that expected from the fitness of the individual mutants (Fig 7). The most obvious example of a genetic interaction is the synthetic lethal, where two individual mutants have no or little fitness effect but abolish growth when combined, and can occur if two gene products perform redundant essential functions. A variety of such interactions exist, which imply different types of relationships between components (reviewed in¹¹⁹). GI networks can be combined with *in vivo* studies to further inform gene function in the host. For example, a gene important for intrinsic resistance to cell-wall targeting antibiotics, that is important in healthy mice but unimportant when certain immune components are missing, indicates that the gene-product is not only a resistance factor but potentially visible to the immune system (Fig 4). Several studies have successfully used a GI approach with TIS in *S. pneumoniae* including to uncover regulatory dependencies for catabolite control protein A; how a potassium uptake system and a sub-pathway for nasopharyngeal colonization are regulated; and how the protease ClpP is involved in competence^{1,35,57}. Additionally, cell division components such as CozE in *S. pneumoniae* were identified by using PBP1A as a query gene, revealing that CozE directs the activity of PBP1a to the mid-cell plane where it promotes zonal cell elongation¹²⁰. Alternatively, a query gene/pathway can be inhibited by a drug or inhibitor, as has been done in the case of wall teichoic acid (WTA) biosynthesis in *S. aureus*, and then screened with TIS for synthetic lethals, which connected WTAs to other pathways, including cell envelope D-alanylation, and peptidoglycan and lipoteichoic acid synthesis¹²¹. Other applications of note include investigating the role of the quorum-sensing and virulence regulator LasR in different *P. aeruginosa* infection models¹²², and assigning function to uncharacterized genes in *M. tuberculosis*.¹²³

Both TIS and RNA-Seq data have shown that even relatively simple perturbations in bacteria (e.g. changes in pH, exposure to antibiotics) trigger complex responses. There is value in combining these data sets; for example, combined TIS and RNA-Seq antibiotic response data obtained from *P. aeruginosa* enabled predictions of antagonistic antibiotic combinations⁵¹ (Fig 7A). Such observations suggest that TIS and RNA-Seq may register distinct, but complementary features of the underlying network architecture of the cell and illustrate how

responses can be separated into at least two organizational levels (phenotypic and transcriptional). Remarkably, when expression and fitness are directly compared (Fig 7B.1) there is often little correlation between them (Fig 7B.2)^{124 125 126}, with some notable exceptions such as in some metabolic pathways¹²⁵ and classical virulence factors⁴¹. This discordance has long been known in yeast^{127 128 129}, and suggests that the majority of transcriptional regulation is not optimized by selection¹³⁰.

Considering RNA-Seq and TIS in the context of the underlying cellular network can clarify their relationship¹²⁴ (Fig 7B.3). When responding to an environmental stimulus a bacterium is adapted to, for instance nutrient depletion, the majority of fitness and expression changes occur at short distances from each other within the metabolic network, with 80% of genes with fitness changes connected by two or fewer metabolic reactions to genes with expression changes, and 93% within a three reaction radius¹²⁴ (Fig 7B.4). Further, the correlation of fitness and expression changes decreases with distance in the network¹²⁴. This indicates that fitness and expression changes are not only co-located within the network, but are of a comparable magnitude to their neighbors. These relationships can disappear when a bacterium responds to an environment it is not adapted to, for instance an antibiotic¹²⁴ (Fig 7B.4). This suggests that the apparently paradoxical lack of correlation between fitness and expression measurements can be in part understood through network models that incorporate regulatory and genetic relationships, which could aid drug target predictions and genetic network engineering. Quantifying the degree of disruption a stimulus creates in a bacterium's transcriptional network has already enabled accurate predictions of fitness, antibiotic sensitivity, and drug mechanism of action¹³¹.

Conclusions and future perspectives

Significant advances and extensions of TIS have been made since its introduction in 2009, many of which are highlighted in this review. Taking advantage of advances in cell sorting and microfluidics, TIS has been adapted to examine phenotypes on the level of single mutant cells^{13 132}. TIS has also demonstrated its practical utility, for example, in the development of new vaccine candidates^{80 79} and new antibiotics or helper drug targets⁴⁴. Lastly, genetic interaction approaches are beginning to build networks that map out the complex relationships between genetic components within the cell, and these could be further extended by combining two transposons into a single genome, by combining TIS and

CRISPR-based transcriptional interference (CRISPRi), or by scaling approaches that simultaneously mutagenize interacting organisms¹³³.

Further, the vast majority of TIS studies to date have been performed in bacterial species, due to this ease of genetic manipulation. For those species that TIS works well in, it is a powerful technique that can provide high volumes of valuable genotypic-phenotype linkage data on a fine scale. Over the next decade, we hope to see an expansion in the types of organisms assayed using TIS methods. Encouragingly, a number of related TIS-like methods have been developed for use in mammalian, fungal, parasite and archaeal backgrounds. Profiles of TIS applied in the two model yeasts, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* identified essential genes, TOR control genes, and factors that contribute to the formation of heterochromatin^{134–136}. Other examples include: determining the essential genes of the archaeal species *Sulfolobus islandicus*¹³⁷ and *Methanococcus maripaludis*¹³⁸; PhITSeq in haploid human cells to assign gene function^{139,140}; QI-Seq in mice using PiggyBac and Sleeping Beauty transposons to screen for cancer related genes^{141,142}; QI-Seq in *Plasmodium falciparum*, using the piggyBac transposon to determine essential genes¹⁴³ and BarSeq in yeast^{144–146}. Moreover, we expect to see the biological questions answered with TIS to become increasingly complex. For this, we must move past the straightforward assays measuring planktonic growth to more sophisticated ones that better mimic natural ecology. Such applications will only enhance the utility and breadth of TIS approaches in the future.

The analogous functional genomics method of pooled CRISPRi screening, which silences genes in a targeted fashion using single guide RNAs (sgRNAs), has been successfully applied to numerous bacterial species after the development of mobile CRISPRi systems^{145–147}. TIS has some distinct advantages over this method, importantly that it can interrogate intergenic regions, including non-coding RNAs, promoter regions and even dissect functional domains within proteins⁴⁰. Additionally, the implementation of TIS requires no prior specific knowledge of the genetic make-up of an organism, whereas sgRNAs need to be predesigned for CRISPRi screens. Further, TIS can assay the effects of gene overexpression and suppression (in the gain of function screens) and also avoids any potential dCas9-specific toxicity or off-target effects. Collectively, these functional genomics screening techniques suffer from similar shortfalls, namely that deciphering detailed mechanistic insight from the large amounts generated can be difficult to automate. For this, research groups will have to pool data, resources and expertise to construct holistic workflows that can manage this. To

this end, we expect to see an increase in data sharing platforms, such as the newly established TIS depository TraDIS-vault (<https://tradis-vault.qfab.org/apollo/jbrowse/>) or interactive data visualisation platforms, like Shinyomics⁴³.

Beyond this, we believe that TIS methods will be applied to answer increasingly complex and diverse biological questions. For this, we must move past the straightforward, planktonically-grown assays to more sophisticated ones that better mimic natural ecological states that occur in nature. One major limitation of TIS remains that it is only available for use in easily culturable and genetically tractable species, which represent only the minority of total bacterial and archaeal species¹⁴⁸. A key challenge will be to develop the tools to allow the recalcitrant microbes of medical, industrial and environmental importance to be assayed. TIS inspired methods, such as the “magic pool” approach to optimizing transposon delivery in non-model organisms²⁸, is already beginning to address this. A driving factor will be massive upscaling in the numbers of conditions and bacterial strains that can be simultaneously screened building on the RB-Tn-Seq²⁷ and similar approaches. As these methods push beyond model strains we will increasingly gain insight into how the genetic diversity within pan-genomes is shaped and maintained. This migration away from one-dimensional genotype versus phenotype experiments will require consideration of the larger genetic network and particularly interactions between genetic background and fitness (see the section “Scaling up TIS using high-throughput phenotyping” above). This will involve the application of machine learning, modelling, and network analyses to integrate and extract knowledge from accumulating TIS datasets. Eventually, and in combination with other post-genomic functional data, this approach will increasingly enable us to move from describing the genetic architecture of the cell to predicting future behaviours^{124 149}. Collectively, these developments illustrate that the journey of TIS is far from over with many exciting paths yet to explore.

Figure Legends

Figure 1: Basic TIS method overview. A | Creating the transposon-insertion sequencing (TIS) library has 4 steps: 1.) Create random mutants: The black horizontal lines and arrows are the host’s genomic DNA (gDNA), the coding regions of genes are marked ‘A, B and C’. The horizontal blue line is the transposon (Tn) containing an antibiotic resistance (Ab^R)

selection marker and bounded by inverted repeats, shown in green. When the Tn inserts into the gDNA, the disruption in a gene (B in this example) is shown by a red cross. 2) Select and pool mutants. The red cross represents a single mutation in each cell, which are selected, for instance on antibiotic-containing agar plates, pooled together and DNA is extracted. 3.) Fragment, add adaptors and PCR amplify. Fragmentation (vertical dashed lines) can be enzymatic or by shearing (depending on the version of TIS). Sequencing adaptors (yellow square) are the added, and primers (purple arrows) P1 and P2 are used to PCR amplify 4.) Sequencing and mapping. Sequences out from the Tn end (using primer P3) are mapped onto the reference genome and the Tn insertion point determined (vertical red arrow), and mapped for each mutant. Genes that cannot tolerate insertion (gene C in this example) will not have any insertions. **B** | Challenging the TIS library — in this example with an antibiotic, colistin (bottom row), compared to an untreated control (top row) (data from REF⁴⁴). The vertical lines denote the density of insertions at each insertion site and red and blue denote forward or reverse insertion direction, respectively. Below are the predicted genes, in turquoise. The first gene (*usq*) has equivalent numbers of insertions in the treated and untreated samples and thus has no effect on fitness in colistin, the next gene (*truA*) has relatively more insertions in the treated sample compared to the control (its mutants have increased fitness in colistin) and thus considered a sensitivity gene, the next gene (*dedA*) is an experimentally confirmed resistance gene⁴⁴ and it has decreased insertions in the treated sample (mutants have decreased fitness in colistin), and the last 2 genes (*accD_1* and *folC*) have no insertions in either treated or untreated and are thus considered essential for growth.

Figure 2: Extensions to the TIS method. **A** | Physical separation of mutant populations based on motility. This includes assaying genes for motility by inoculating on an agar plate (yellow circle and beige spot) and separating the inner (less motile) from the outer (more motile) mutant pools. **B** | In density-TraDISort, mutant populations can be separated into top, mid and bottom fractions (shown by orange horizontal bands) based on their increased or decreased cellular density (density-TraDISort), using a Percoll gradient and centrifugation. **C** | Separation of single mutants using fluorescence (TraDISort). The mutant pool is treated with the fluorescent ethidium bromide (EthBr) and subjected to fluorescence-activated cell sorting (FACS) where each cell is sorted via laser (red horizontal line) based on its fluorescence (shown as green), reporting on efflux activity. **D** | Encapsulation, growth and sorting with microfluidics of single mutants in droplets for droplet Tn-Seq. Each single

mutant with different growth rates (in the schematic on the left: low, mid and high levels of growth are represented as blue, orange and green background colors, respectively) will grow independently within its own droplet (grey circle), eliminating the effects of interactions between mutants. A final sorting step, based on cell fluorescence or microscopy can also be added. Alternatively, cell-containing droplets (blue droplet on the right) can undergo multiple layers of re-encapsulation, so that an encapsulated mutant can be encapsulated within another droplet containing a different cell (shown in orange; this can be another mutant, another bacterial cell or host cell) and signals can freely diffuse between the layers (shown as a red bolt) to allow cell interactions to be investigated by sorting those combinations that have altered fitness and grow at different rates, or those that can be separated by sorting based on markers, such as alterations of cell morphology using a microscope.

Figure 3: TIS to assay function of essential genes. A) An inducible promoter (red, right-angled arrow), such as pBAD, is positioned facing out of each transposon (Tn) to overexpress (orange reads for induced transcription) all genes, including essential genes, so that their function can be assayed. B) In traditional transposon insertion sequencing approaches, essential genes can be identified as those that cannot tolerate insertions (top panel). In gain of function screens when Tns in the transposon pool are induced, for example with arabinose (second panel), high expression of essential gene C is achieved. When selection is applied, involvement of essential genes in the condition, can now be assayed (bottom panel) via monitoring relative differences in the number of transposons that influence expression levels. In this example, after exposure to the condition and sequencing out of the Tn, an increase in the number of Tn insertions that increase gene C's expression is observed. Therefore mutants that overexpress gene C have increased fitness within the overall population during selection indicated that C's expression is beneficial in that condition. All other features are as in Figure 1.

Figure 4. Mapping complex genotype–phenotype relationships. A.1 | A gene–antibiotic network for 3 antibiotics (based on ⁵⁷ and unpublished data from the van Opijnen laboratory). Each node (circle) depicts a gene while each edge indicates a negative (black), neutral (dashed) or positive (red) effect on fitness between the gene in the presence of antibiotics as determined using transposon-insertion sequencing (TIS). All 3 antibiotics, penicillin (PeniG); vancomycin (Vanco); daptomycin (Dapto), affect cell wall integrity, but TIS uncovers a wide

variety of genes involved many of them not direct targets of the antibiotic. In this case, the unknown gene Q is likely to be involved in peptidoglycan synthesis/membrane integrity, based on the function of other genes with similar fitness profiles. By mutating gene Q and constructing a Tn-library in this mutant background, all genetic interactions can be identified. In this case, the genes uncovered in the mutant background, as depicted in the Gene Interaction Map (GIM), further support a role in peptidoglycan synthesis (A.2), that it may function in the cell membrane, and may be controlled by a particular regulator. Additionally, *in vivo* TIS data with the mutant library performed in healthy mice and mice depleted for neutrophils (neut-), indicates that gene Q is needed to establish lung infection, but dispensable in the absence of neutrophils.

Figure 5: Bottleneck and realism trade-off in TIS infection models. Different animal models of infection (top series) can induce strong bottleneck effects (bottom series) in infecting bacterial populations, which can confound transposon-insertion sequencing (TIS) analysis. This is particularly pronounced in models where bacteria must overcome barrier defences (e.g. oral infection models of *Salmonella* Typhimurium). Organoid models can provide a complex environment while limiting bottlenecks, but technical limitations in culture may limit the total population size screened. Finally, cell culture screens can be scaled to arbitrary sizes, allowing for screening of extremely large collections of mutants, but they often only provide insight into a particular aspect of disease.

Figure 6: TIS to assay microbe-microbe interactions. A | Transposon-insertion sequencing (TIS) of *Staphylococcus aureus* (SA) mutants alone (golden circles) compared to co-culture of SA and *Pseudomonas aeruginosa* (PA) wild type strain (green rods) identifying community dependent essential genes that are needed only during co-culture. Other features are as in Fig 1. **B** | Using TIS to identify type 6 secretion system (T6SS) toxin immunity pairs by growing cells with close cell-to-cell contact, and performing TIS so genes involved in T6SS-depending killing can be detected. In this example, genes in *P. aeruginosa* (PA) for immunity proteins (yellow diamond) that protect from neighbour killing by toxins (red diamond) become essential only in a T6SS active (*retS*; bottom panel) background, but not in the inactive (*H1_retS*; top panel) cellular background.

Figure 7. Integrating TIS with RNA-Seq data.

A | An example of combining RNA-Seq (depicted in green throughout) and TIS (depicted in red throughout), to identify antagonistic interactions between antibiotics polymyxin B (PolyB) and gentamicin (Gent) or tobramycin (Tobr) in *P. aeruginosa*. Tn-Seq data indicates that genes MexYX are involved in intrinsic resistance in *P. aeruginosa* to gentamicin and tobramycin, indicated by the red edges between the antibiotics and the genes. When these genes are disrupted by a Tn-insertion, the bacterium becomes more sensitive to these antibiotics. Moreover, RNA-Seq data indicates that polymyxin B induces expression of these genes, as indicated by the green arrows. This led to the hypothesis that polymyxin B, due to its transcriptional activation of MexYX, will make the bacterium less sensitive to either gentamicin or tobramycin. The authors confirmed experimentally that these antibiotics work in an antagonistic fashion⁵¹, which highlights the strength of probing response networks from different perspectives to extract biological meaning.

B | Measuring TIS and RNA-Seq responses under the same conditions (panel B.1) has shown that the transcriptional responses to a specific environment (Δ expression) are often not accurate predictors of gene deletion phenotype, as expression and fitness do not correlate well (Δ fitness; B.2)¹²⁴. However, by overlaying these datasets over a known network (e.g. a metabolic network; B.3), network analyses can identify patterns between Tn-Seq and RNA-Seq. In this example a small part of a metabolic network is depicted; grey circles are metabolites, arrows are genes encoding enzymes that mediate each reaction. Red arrows are phenotypically important genes in a specific environment identified with TIS, while green arrows are genes that change transcriptionally in the same environment, identified by RNA-Seq. The distance between two genes in a metabolic network is the number of reactions between them, and can be calculated for all pairs across the network. In B.4, the top network is an example where distances between pairs of fitness and expression changes are small, while the bottom network illustrates larger distances. An adapted response to an environment is characterized by fitness and expression changes that are relatively small (distance to neighbour) and correlated (Δ fitness x expression). Exposure to stress conditions a bacterium is not adapted to leads to a loss in correlation between genes that change in transcription and those have a fitness effect (B.4)¹²⁴. Importantly, such associations can be used to make predictions on antibiotic susceptibility¹³¹.

772 **Table 1:** Transposon-insertion sequencing data analysis tools.

Name	Interface*	Unique features	Tn	URL	REF
INSeq pipeline	CL	Basic mapping pipeline	Any	https://www.nature.com/articles/nprot.2011.417	106
ESSENTIALS	Web	Essentiality and conditional analysis, corrects for genomic location biases	Any	http://bamics2.cmbi.ru.nl/websites/essentials/essentials_start.php	29
ARTIST/conARTIST	CL	HMM-based mapping for fitness loci, sampling based bottleneck correction	Mariner	https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1004782#s4	32
TRANSIT	GUI	Bayesian, HMM, and resampling-based tests	Mariner	https://orca1.tamu.edu/essentiality/transit/	31
TraDIS Toolkit	CL	Empirical essentiality analysis, comparative analysis with RNA-seq tools	Any	https://github.com/sanger-pathogens/Bio-TraDis	33
MAGenTA	CL or Web	Fitness as growth rate calculations. Bottleneck correction using neutral loci, sliding window analyses	Any	https://vanopijnenlab.github.io/MAGenTA/	30
Aerobio	CL	Fitness as growth rate calculations. Bottleneck correction using neutral loci, sliding window analyses. Ability to switch to RNA-seq, Term-Seq and WGS analyses.	Any	https://github.com/jsa-aerial/aerobio	34
AlbaTraDIS	CL	Multicondition and gene activation analysis based on the TraDIS Toolkit	Any	https://github.com/quadrant-institute-bioscience/albatradis	40

*GUI, standalone graphical user interface; CL, command-line (*NIX-based) utility; Web, web-based graphical user interface

775 ***Glossary***

776

777 TRANPOSON

778 A mobile genetic element that inserts itself into a bacterial genome and disrupts genes or
779 genetic features at that site.

780

781 GENETIC FEATURES

782 Every component of the genome that can be annotated as being a feature, whether it be a
783 gene, coding RNA, non-coding RNA or promoter region.

784

785 NEXT-GENERATION SEQUENCING

786 Genetic sequences generated using a massively parallel platform that separates DNA
787 templates on a flow cell and clonal amplifies clusters for sequencing.

788

789 FRAGMENTATION

790 Breaking up DNA into smaller pieces in order to be sequenced. This can be done using
791 physical shearing methods, such as sonication, or enzymatic digestion

792

793 INTRINSIC ANTIBIOTIC RESISTANCE GENES

794 Genetic components of the bacterial genome that, when present and/or active, allow the
795 bacteria to tolerate exposure to antibiotics that would otherwise kill them, or inhibit their
796 growth.

797

798 FACS

799 Fluorescence-activated cell sorting is a specialised type of the flow cytometry method that
800 separates cells, one cell at time, by their fluorescent characteristics based on light scattering.

801

802 SLIDING WINDOW

803 A window of arbitrary length is set and events within that window are assayed and this
804 window is then moved around the genome, thus providing a more objective method of
805 assessing the genome that is independent of annotation.

806

807 BOTTLENECK

808 When a population is drastically reduced through stochastic processes, and the surviving cells
809 will make up the new population, but will have reduced genetic diversity

810

811 ESSENTIAL GENOME

812 The complete set of genes and genetic features in a genome that are essential for a cell to
813 survive and grow, the exemplar of which are “housekeeping” genes for core processes such
814 as replication and division.

815

816 GNOTOBIOTIC

817 An environment for culturing microorganisms, such as an animal model, where all microbes
818 are either defined or removed

819

820 PAN-GENOME

821 The complete set of genes in all bacterial strains within a species. Contrast to the core
822 genome: the set of genes shared by all strains within a species.

823

824 OVERDISPERSION

825 When data exhibits greater variability than would be expected under a statistical model. In
826 the context of sequencing data, often used in reference to the negative binomial distribution
827 which can be understood as a generalization of the Poisson distribution that allows for a
828 larger variance relative to the mean.

829

830 ORGANOID

831 A 3-dimensional, simplified replica of an organ derived from stem cells to realistically model
832 the organ *in-vitro*

833

834 REVERSE GENETICS

835 Determining the phenotypic effects of a genetic feature, by either altering the genetic feature
836 and observing changes in the organism compared to a wild-type.

837

838 ANTAGONISTIC ANTIBIOTIC COMBINATION

839 When the activity of an antibiotic combination is lower than would be predicted from the
840 individual antibiotics’ effects

841

842 SYNTHETIC LETHAL

843 Where individual mutants have no or little fitness effect but when two or more mutations are
844 combined, this leads to arrest in cell growth or death

845

References

1. 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).
2. Langridge, G. C. *et al.* Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. *Genome Res.* **19**, 2308–2316 (2009).
3. 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).
4. Gawronski, J. D. *et al.* Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for Haemophilus genes required in the lung. *Proc. Natl. Acad. Sci.* **106**, 16422–16427 (2009).
5. Van Opijnen, T. & Camilli, A. Transposon insertion sequencing: A new tool for systems-level analysis of microorganisms. *Nat. Rev. Microbiol.* **11**, 435–442 (2013).

A previous review on TIS that outlines the detail and difference of each variation of TIS

6. Barquist, L., Boinett, C. J. & Cain, A. K. Approaches to querying bacterial genomes with transposon-insertion sequencing. *RNA Biol.* **10**, 1161–1169 (2013).

A previous review on TIS that outlines the use and all applications of TIS until 2013

7. Chao, M. C., Abel, S., Davis, B. M. & Waldor, M. K. The design and analysis of transposon insertion sequencing experiments. *Nat. Rev. Microbiol.* **14**, 119–128 (2016).
8. Van Opijnen, T. & Camilli, A. Transposon insertion sequencing: A new tool for systems-level analysis of microorganisms. *Nat. Rev. Microbiol.* **11**, 435–442 (2013).
9. Barquist, L., Boinett, C. J. & Cain, A. K. Approaches to querying bacterial genomes with transposon-insertion sequencing. *RNA Biology* **10**, 1161–1169 (2013).
10. Kakkanat, A., Phan, M.-D., Lo, A. W., Beatson, S. A. & Schembri, M. A. Novel genes associated with enhanced motility of Escherichia coli ST131. *PLoS One* **12**, e0176290 (2017).
11. Nolan, L. M. *et al.* A global genomic approach uncovers novel components for twitching motility-mediated biofilm expansion in Pseudomonas aeruginosa. *Microb. Genomics* **4**, (2018).

One of the first studies to apply physical separation methods to TIS, in this case to examine twitching motility based biofilm formation

12. 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**,

(2018).

13. 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**, (2016).

One of the first methods to separate single mutant cells using FACS and sequence them using TIS, in this case to examine efflux of each cell of the biocide ethidium bromide

14. 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).
15. Fowler, C. C. & Galán, J. E. Decoding a *Salmonella Typhi* Regulatory Network that Controls Typhoid Toxin Expression within Human Cells. *Cell Host Microbe* **23**, 65–76.e6 (2018).
16. Duncan, M. C. *et al.* High-throughput analysis of gene function in the bacterial predator *Bdellovibrio bacteriovorus*. *MBio* **10**, 1–12 (2019).
17. Thibault, D. *et al.* Droplet Tn-Seq combines microfluidics with Tn-Seq for identifying complex single-cell phenotypes. *Nat. Commun.* **10**, 5729 (2019).

The first method to separate and individually encapsulate single cells, in order to assay each mutant's effects independently from the population

18. Judson, N. & Mekalanos, J. J. TnAraOut, A transposon-based approach to identify and characterize essential bacterial genes. *Nat. Biotechnol.* **18**, 740–745 (2000).
19. Christen, B. *et al.* The essential genome of a bacterium. *Mol. Syst. Biol.* **7**, 1–7 (2011).
20. Coe, K. A. *et al.* Multi-strain Tn-Seq reveals common daptomycin resistance determinants in *Staphylococcus aureus*. *PLOS Pathog.* **15**, e1007862 (2019).
21. Santiago, M. *et al.* A new platform for ultra-high density *Staphylococcus aureus* transposon libraries. *BMC Genomics* **16**, 1–18 (2015).
22. Yasir, M. *et al.* TraDIS-Xpress: a high-resolution whole-genome assay identifies novel mechanisms of triclosan action and resistance. *Genome Res.* **30**, 1–11 (2020).
23. Yasir, M. *et al.* A new massively-parallel transposon mutagenesis approach comparing multiple datasets identifies novel mechanisms of action and resistance to triclosan. *bioRxiv* (2019).
24. Mutalik, V. K. *et al.* Dual-barcoded shotgun expression library sequencing for high-throughput characterization of functional traits in bacteria. *Nat. Commun.* **10**, 308 (2019).
25. Wetmore, K. M. *et al.* Sequencing Randomly Bar-Coded Transposons. *MBio* **6**, 1–15

(2015).

26. Robinson, D. G., Chen, W., Storey, J. D. & Gresham, D. Design and Analysis of Bar-seq Experiments. *G3 (Bethesda, Md.)*; *Genes|Genomes|Genetics* **4**, 11–18 (2014).
27. Price, M. N. *et al.* Mutant phenotypes for thousands of bacterial genes of unknown function. *Nature* **557**, 503–509 (2018).

A massively upscaled study that applied TIS to multiple bacterial species and over 100 conditions in order to assign gene broader gene function *en masse*.

28. Liu, H. *et al.* Magic Pools: Parallel Assessment of Transposon Delivery Vectors in Bacteria. *mSystems* **3**, 1–17 (2018).
29. Zomer, A., Burghout, P., Bootsma, H. J., Hermans, P. W. M. & van Hijum, S. A. F. T. Essentials: Software for rapid analysis of high throughput transposon insertion sequencing data. *PLoS One* **7**, 1–9 (2012).
30. McCoy, K. M., Antonio, M. L. & van Opijnen, T. MAGenTA: a Galaxy implemented tool for complete Tn-Seq analysis and data visualization. *Bioinformatics* **33**, 2781–2783 (2017).
31. DeJesus, M. A., Ambadipudi, C., Baker, R., Sassetti, C. & Ioerger, T. R. TRANSIT - A Software Tool for Himar1 TnSeq Analysis. *PLoS Comput. Biol.* **11**, 1–17 (2015).
32. Pritchard, J. R. *et al.* ARTIST: High-Resolution Genome-Wide Assessment of Fitness Using Transposon-Insertion Sequencing. *PLoS Genet.* **10**, (2014).
33. Barquist, L. *et al.* The TraDIS toolkit: sequencing and analysis for dense transposon mutant libraries. *Bioinformatics* **32**, 1109–1111 (2016).
34. Anthony, J. S. and T. van O. A DAG computation server for fully integrated and automated massively parallel sequencing analyses. *GitHub* (2019). Available at: <https://github.com/jsa-aerial/aerobio>.
35. Van Opijnen, T. & Camilli, A. A fine scale phenotype-genotype virulence map of a bacterial pathogen. *Genome Res.* **22**, 2541–2551 (2012).
36. Warr, A. R. *et al.* Transposon-insertion sequencing screens unveil requirements for EHEC growth and intestinal colonization. *PLoS Pathogens* **15**, (2019).
37. Subramaniyam, S. *et al.* Statistical analysis of variability in TnSeq data across conditions using zero-inflated negative binomial regression. *BMC Bioinformatics* **20**, 603 (2019).
38. Van den Berge, K. *et al.* Observation weights unlock bulk RNA-seq tools for zero inflation and single-cell applications. *Genome Biol.* **19**, 24 (2018).
39. Hubbard, T. P., D’Gama, J. D., Billings, G., Davis, B. M. & Waldor, M. K.

- Unsupervised Learning Approach for Comparing Multiple Transposon Insertion Sequencing Studies. *mSphere* **4**, (2019).
40. Page, A. J. *et al.* AlbaTraDIS: Comparative analysis of large datasets from parallel transposon mutagenesis experiments. *bioRxiv* 593624 (2019). doi:10.1101/593624
 41. Canals, R. *et al.* The fitness landscape of the African Salmonella Typhimurium ST313 strain D23580 reveals unique properties of the pBT1 plasmid. *PLOS Pathog.* **15**, e1007948 (2019).
 42. Down, T. A., Piipari, M. & Hubbard, T. J. P. Dalliance: Interactive genome viewing on the web. *Bioinformatics* **27**, 889–890 (2011).
 43. T., S. D. and van O. ShinyOmics: Collaborative online exploration, sharing and management of omics-data. *BMC Bioinformatics*
 44. Jana, B. *et al.* The secondary resistome of multidrug-resistant *Klebsiella pneumoniae*. *Sci. Rep.* (2017). doi:10.1038/srep42483
 45. Boinett, C. J. *et al.* Clinical and laboratory-induced colistin-resistance mechanisms in *Acinetobacter baumannii*. *Microb. genomics* **5**, (2019).
 46. Gallagher, L. A., Lee, S. A. & Manoil, C. Importance of Core Genome Functions for an Extreme Antibiotic Resistance Trait. *MBio* **8**, (2017).
 47. Coe, K. A. *et al.* Comparative Tn-Seq reveals common daptomycin resistance determinants in *Staphylococcus aureus* despite strain-dependent differences in essentiality of shared cell envelope genes. *bioRxiv* (2019).
 48. Geisinger, E., Mortman, N. J., Vargas-Cuevas, G., Tai, A. K. & Isberg, R. R. A global regulatory system links virulence and antibiotic resistance to envelope homeostasis in *Acinetobacter baumannii*. *PLOS Pathog.* **14**, e1007030 (2018).
 49. Xu, W. *et al.* Chemical Genetic Interaction Profiling Reveals Determinants of Intrinsic Antibiotic Resistance in *Mycobacterium tuberculosis*; *Antimicrob. Agents Chemother.* **61**, e01334-17 (2017).
 50. Willcocks, S. *et al.* Genome-wide assessment of antimicrobial tolerance in *Yersinia pseudotuberculosis* under ciprofloxacin stress. *Microb. Genomics* (2019). doi:10.1099/mgen.0.000304
 51. Murray, J. L., Kwon, T., Marcotte, E. M. & Whiteley, M. Intrinsic antimicrobial resistance determinants in the superbug *Pseudomonas aeruginosa*. *MBio* **6**, 1–10 (2015).
 52. Rajagopal, M. *et al.* Multidrug Intrinsic Resistance Factors in *Staphylococcus aureus* Identified by Profiling Fitness within High-Diversity Transposon Libraries. *MBio* **7**,

- (2016).
53. Blake, K. L. & O'Neill, A. J. Transposon library screening for identification of genetic loci participating in intrinsic susceptibility and acquired resistance to antistaphylococcal agents. *J. Antimicrob. Chemother.* **68**, 12–16 (2012).
 54. Gallagher, L. A., Shendure, J. & Manoil, C. Genome-scale identification of resistance functions in *Pseudomonas aeruginosa* using Tn-seq. *MBio* **2**, (2011).
 55. Cain, A. K. *et al.* Transposon Directed Insertion-Site Sequencing (TraDIS) to Elucidate the Mode of Action of the Antimicrobial Arenicin-3 (Arn-3). in *54th International Conference of Antimicrobial Agents and Chemotherapy (ICAAC) and Infectious Diseases Society of America (IDSA)* (2014).
 56. Santiago, M. *et al.* Genome-wide mutant profiling predicts the mechanism of a Lipid II binding antibiotic. *Nat. Chem. Biol.* **14**, 601–608 (2018).
 57. van Opijnen, T., Dedrick, S. & Bento, J. Strain Dependent Genetic Networks for Antibiotic-Sensitivity in a Bacterial Pathogen with a Large Pan-Genome. *PLoS Pathog.* **12**, e1005869 (2016).
 58. Fields, P. I., Swanson, R. V., Haidaris, C. G. & Heffron, F. Mutants of *Salmonella typhimurium* that cannot survive within the macrophage are avirulent. *Proc. Natl. Acad. Sci. U. S. A.* **83**, 5189–5193 (1986).
 59. Hensel, M. *et al.* Simultaneous identification of bacterial virulence genes by negative selection. *Science* (80-.). **269**, 400 LP – 403 (1995).
 60. Goh, K. G. K. *et al.* Genome-Wide Discovery of Genes Required for Capsule Production by Uropathogenic *Escherichia coli*. *MBio* **8**, (2017).
 61. McCarthy, A. J., Stabler, R. A. & Taylor, P. W. Genome-Wide Identification by Transposon Insertion Sequencing of *Escherichia coli* K1 Genes Essential for In Vitro Growth, Gastrointestinal Colonizing Capacity, and Survival in Serum. *J. Bacteriol.* **200**, (2018).
 62. Sharp, C. *et al.* O-Antigen-Dependent Colicin Insensitivity of Uropathogenic *Escherichia coli*. (2019). doi:10.1128/JB
 63. Dembek, M. *et al.* High-Throughput Analysis of Gene Essentiality and Sporulation in *Clostridium difficile*. **6**, 1–13 (2015).
 64. Chaudhuri, R. R. *et al.* Comprehensive Assignment of Roles for *Salmonella Typhimurium* Genes in Intestinal Colonization of Food-Producing Animals. *PLoS Genet.* **9**, (2013).
 65. Vohra, P. *et al.* Retrospective application of transposon-directed insertion-site

- sequencing to investigate niche-specific virulence of Salmonella Typhimurium in cattle. *BMC Genomics* **20**, 20 (2019).
66. Shames, S. R. *et al.* Multiple Legionella pneumophila effector virulence phenotypes revealed through high-throughput analysis of targeted mutant libraries. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E10446–E10454 (2017).
 67. Zhu, L. *et al.* Novel Genes Required for the Fitness of Streptococcus pyogenes in Human Saliva. *mSphere* **2**, (2017).
 68. Subashchandrabose, S. *et al.* Acinetobacter baumannii Genes Required for Bacterial Survival during Bloodstream Infection. *mSphere* **1**, (2016).
 69. AS Charbonneau ARL; Taylor E; Mitchell CJ; Robinson C; Cain AK; Leigh JA; Maskell DJ; Waller. Identification of genes required for the fitness of Streptococcus equi subsp. equi in whole equine blood and hydrogen peroxide. *Microb. Genomics* **accepted**, (2020).
 70. Wang, H. *et al.* Hypermutation-induced in vivo oxidative stress resistance enhances Vibrio cholerae host adaptation. *PLoS Pathog.* **14**, e1007413 (2018).
 71. Wong, Y.-C. *et al.* Genetic Determinants Associated With in Vivo Survival of Burkholderia cenocepacia in the Caenorhabditis elegans Model. *Front. Microbiol.* **9**, 1118 (2018).
 72. Shields, R. C., Zeng, L., Culp, D. J. & Burne, R. A. Genomewide Identification of Essential Genes and Fitness Determinants of Streptococcus mutans UA159. *mSphere* **3**, (2018).
 73. Crabill, E., Schofield, W. B., Newton, H. J., Goodman, A. L. & Roy, C. R. Dot/Icm-Translocated Proteins Important for Biogenesis of the Coxiella burnetii-Containing Vacuole Identified by Screening of an Effector Mutant Sublibrary. *Infect. Immun.* **86**, (2018).
 74. Grant, A. J. *et al.* Modelling within-Host Spatiotemporal Dynamics of Invasive Bacterial Disease. *PLOS Biol.* **6**, e74 (2008).
 75. Abel, S., Abel zur Wiesch, P., Davis, B. M. & Waldor, M. K. Analysis of Bottlenecks in Experimental Models of Infection. *PLoS Pathog.* **11**, e1004823–e1004823 (2015).
 76. Carter, P. B. & Collins, F. M. THE ROUTE OF ENTERIC INFECTION IN NORMAL MICE . *J. Exp. Med.* **139**, 1189–1203 (1974).
 77. Maier, L. *et al.* Granulocytes Impose a Tight Bottleneck upon the Gut Luminal Pathogen Population during Salmonella Typhimurium Colitis. *PLOS Pathog.* **10**, e1004557 (2014).

78. Lawley, T. D. *et al.* Genome-wide screen for Salmonella genes required for long-term systemic infection of the mouse. *PLoS Pathog.* **2**, e11–e11 (2006).
79. Rowe, H. M. *et al.* Bacterial Factors Required for Transmission of Streptococcus pneumoniae in Mammalian Hosts. *Cell Host Microbe* **25**, 884-891.e6 (2019).
80. Carter, R. *et al.* Genomic analyses of pneumococci from children with sickle cell disease expose host-specific bacterial adaptations and deficits in current interventions. *Cell Host Microbe* **15**, 587–599 (2014).
81. Hancock, S. J. *et al.* Identification of IncA/C Plasmid Replication and Maintenance Genes and Development of a Plasmid Multilocus Sequence Typing Scheme. *Antimicrob. Agents Chemother.* **61**, (2017).
82. Liu, Y. *et al.* Transposon insertion sequencing reveals T4SS as the major genetic trait for conjugation transfer of multi-drug resistance pEIB202 from Edwardsiella. *BMC Microbiol.* **17**, 1–15 (2017).
83. Cowley, L. A. *et al.* Transposon Insertion Sequencing Elucidates Novel Gene Involvement in Susceptibility and Resistance to Phages T4 and T7 in Escherichia coli O157. *MBio* **9**, (2018).
84. Pickard, D. *et al.* A Genomewide Mutagenesis Screen Identifies Multiple Genes Contributing to Vi Capsular Expression in Salmonella enterica; Serovar Typhi. *J. Bacteriol.* **195**, 1320 LP – 1326 (2013).
85. Klein, B. A., Duncan, M. J. & Hu, L. T. Defining essential genes and identifying virulence factors of Porphyromonas gingivalis by massively parallel sequencing of transposon libraries (Tn-seq). *Methods Mol. Biol.* **1279**, 25–43 (2015).
86. Wong, Y.-C. *et al.* Candidate Essential Genes in Burkholderia cenocepacia J2315 Identified by Genome-Wide TraDIS. *Front. Microbiol.* **7**, 1288 (2016).
87. Willcocks, S. J., Stabler, R. A., Atkins, H. S., Oyston, P. F. & Wren, B. W. High-throughput analysis of Yersinia pseudotuberculosis gene essentiality in optimised in vitro conditions, and implications for the speciation of Yersinia pestis. *BMC Microbiol.* **18**, 46 (2018).
88. Charbonneau, A. R. L. *et al.* Defining the ABC of gene essentiality in streptococci. *BMC Genomics* (2017). doi:10.1186/s12864-017-3794-3
89. McCann, H. C. *et al.* Genomic Analysis of the Kiwifruit Pathogen Pseudomonas syringae pv. actinidiae Provides Insight into the Origins of an Emergent Plant Disease. *PLoS Pathog.* **9**, (2013).
90. Goodall, E. C. A. *et al.* The Essential Genome of Escherichia coli K-12. *MBio* **9**,

(2018).

Using TIS the essential genome of the most well studied bacterial model strain, *E. coli* K12, was defined in detail.

91. Ruiz, L. *et al.* The essential genomic landscape of the commensal *Bifidobacterium breve* UCC2003. *Sci. Rep.* **7**, 5648 (2017).
92. 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).
93. Poulsen, B. E. *et al.* Defining the core essential genome of *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 10072–10080 (2019).
94. Geisinger, E. *et al.* The Landscape of Phenotypic and Transcriptional Responses to Ciprofloxacin in *Acinetobacter baumannii*: Acquired Resistance Alleles Modulate Drug-Induced SOS Response and Prophage Replication. *MBio* **10**, 1–19 (2019).
95. Carey, A. F. *et al.* TnSeq of *Mycobacterium tuberculosis* clinical isolates reveals strain-specific antibiotic liabilities. *PLoS Pathog.* **14**, e1006939–e1006939 (2018).
96. Deutschbauer, A. *et al.* Towards an Informative Mutant Phenotype for Every Bacterial Gene. *J. Bacteriol.* **196**, 3643 LP – 3655 (2014).
97. Basta, D. W., Bergkessel, M. & Newman, D. K. Identification of Fitness Determinants during Energy-Limited Growth Arrest in *Pseudomonas aeruginosa*. *MBio* **8**, (2017).
98. Price, M. N. *et al.* Filling gaps in bacterial amino acid biosynthesis pathways with high-throughput genetics. *PLoS Genet.* **14**, 1–23 (2018).
99. Byrne, R. T., Chen, S. H., Wood, E. A., Cabot, E. L. & Cox, M. M. *Escherichia coli* genes and pathways involved in surviving extreme exposure to ionizing radiation. *J. Bacteriol.* **196**, 3534–3545 (2014).
100. Kingsley, R. A. *et al.* Functional analysis of *Salmonella* Typhi adaptation to survival in water. *Environ. Microbiol.* **20**, 4079–4090 (2018).
101. Mandal, R. K. & Kwon, Y. M. Global Screening of *Salmonella enterica* Serovar Typhimurium Genes for Desiccation Survival. *Front. Microbiol.* **8**, 1723 (2017).
102. Fabian, B. K., Tetu, S. G. & Paulsen, I. T. Application of transposon insertion sequencing to agricultural science. *Front. Plant Sci.* **11**, 291 (2020).
103. Royet, K., Parisot, N., Rodrigue, A., Gueguen, E. & Condemine, G. Identification by Tn-seq of *Dickeya dadantii* genes required for survival in chicory plants. *Mol. Plant Pathol.* **20**, 287–306 (2019).
104. Duong, D. A., Jensen, R. V & Stevens, A. M. Discovery of *Pantoea stewartii* ssp.

stewartii genes important for survival in corn xylem through a Tn-Seq analysis. *Mol. Plant Pathol.* (2018). doi:10.1111/mpp.12669

105. Liu, H. *et al.* Large-scale chemical-genetics of the human gut bacterium *Bacteroides thetaiotaomicron*. *bioRxiv* (2019).
106. 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).
107. Zimmermann, M., Zimmermann-Kogadeeva, M., Wegmann, R. & Goodman, A. L. Separating host and microbiome contributions to drug pharmacokinetics and toxicity. *Science* (80-.). **363**, eaat9931 (2019).

An applied study that has used a TIS arrayed approach to dissect complex host-microbiome interactions during response to drug treatment

108. Stacy, A., Fleming, D., Lamont, R. J., Rumbaugh, K. P. & Whiteley, M. A Commensal Bacterium Promotes Virulence of an Opportunistic Pathogen via Cross-Respiration. *MBio* **7**, (2016).
109. Ibberson, C. B. *et al.* Co-infecting microbes dramatically alter pathogen gene essentiality during polymicrobial infection. *Nat. Microbiol.* 1234- (2018). doi:10.1038/nmicrobiol.2017.79

One of the first studies that have used TIS in a co-infection model, in this case 2 key pathogens in a wound infection

110. Duncan, M. C. *et al.* *Vibrio cholerae* motility exerts drag force to impede attack by the bacterial predator *Bdellovibrio bacteriovorus*. *Nat. Commun.* **9**, 4757 (2018).
111. Duncan, M. C. *et al.* High-Throughput Analysis of Gene Function in the Bacterial Predator *Bdellovibrio bacteriovorus*. *MBio* **10**, (2019).
112. Chatterjee, A. *et al.* Parallel Genomics Uncover Novel Enterococcal-Bacteriophage Interactions. *MBio* **11**, e03120-19 (2020).
113. Fu, Y., Waldor, M. K. & Mekalanos, J. J. Tn-seq analysis of *vibrio cholerae* intestinal colonization reveals a role for T6SS-mediated antibacterial activity in the host. *Cell Host Microbe* **14**, 652–663 (2013).
114. Nolan, L. M. *et al.* Discovery of a *Pseudomonas aeruginosa* Type VI secretion system toxin targeting bacterial protein synthesis using a global genomics approach. *bioRxiv* 733030 (2019). doi:10.1101/733030
115. Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R. & Hofacker, I. L. The Vienna RNA websuite. *Nucleic Acids Res.* **36**, 70–74 (2008).

116. Mann, B. *et al.* Control of virulence by small RNAs in *Streptococcus pneumoniae*. *PLoS Pathog.* **8**, 34 (2012).

One of the first studies to combine RNA-seq and TIS to investigate sRNAs in *S.pneumoniae*

117. Warriar, I. *et al.* The Transcriptional landscape of *Streptococcus pneumoniae* TIGR4 reveals a complex operon architecture and abundant riboregulation critical for growth and virulence. *PLoS Pathog.* **14**, 1–25 (2018).
118. Capel, E. *et al.* Comprehensive Identification of Meningococcal Genes and Small Noncoding RNAs Required for Host Cell Colonization. *MBio* **7**, (2016).
119. Costanzo, M. *et al.* Global Genetic Networks and the Genotype-to-Phenotype Relationship. *Cell* **177**, 85–100 (2019).
120. Fenton, A. K., El Mortaji, L., Lau, D. T. C., Rudner, D. Z. & Bernhardt, T. G. CozE is a member of the MreCD complex that directs cell elongation in *Streptococcus pneumoniae*. *Nat. Microbiol.* **2**, 16237 (2016).
121. Santa Maria, J. P. J. *et al.* Compound-gene interaction mapping reveals distinct roles for *Staphylococcus aureus* teichoic acids. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 12510–12515 (2014).
122. Lorenz, A. *et al.* Importance of flagella in acute and chronic *Pseudomonas aeruginosa* infections. *Environ. Microbiol.* **21**, 883–897 (2019).
123. DeJesus, M. A. *et al.* Statistical analysis of genetic interactions in Tn-Seq data. *Nucleic Acids Res.* **45**, e93 (2017).
124. Jensen, P. A., Zhu, Z. & van Opijnen, T. Antibiotics Disrupt Coordination between Transcriptional and Phenotypic Stress Responses in Pathogenic Bacteria. *Cell Rep.* **20**, 1705–1716 (2017).
125. Turner, K. H., Everett, J., Trivedi, U., Rumbaugh, K. P. & Whiteley, M. Requirements for *Pseudomonas aeruginosa* acute burn and chronic surgical wound infection. *PLoS Genet.* **10**, e1004518 (2014).
126. Deutschbauer, A. *et al.* Evidence-Based Annotation of Gene Function in *Shewanella oneidensis* MR-1 Using Genome-Wide Fitness Profiling across 121 Conditions. *PLOS Genet.* **7**, e1002385 (2011).
127. Smith, J. J. *et al.* Expression and functional profiling reveal distinct gene classes involved in fatty acid metabolism. *Mol. Syst. Biol.* **2**, 2006.0009–2006.0009 (2006).
128. Giaever, G. *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387–391 (2002).

129. Keren, L. *et al.* Massively Parallel Interrogation of the Effects of Gene Expression Levels on Fitness. *Cell* **166**, 1282-1294.e18 (2016).
130. Price, M. N. *et al.* Indirect and suboptimal control of gene expression is widespread in bacteria. *Mol. Syst. Biol.* **9**, 660 (2013).
131. Zeyu Zhu, Defne Surujon, Juan C. Ortiz-Marquez, Stephen J. Wood, Wenwen Huo, Ralph R. Isberg, José Bento, T. van O. Entropy of a bacterial stress response is a generalizable predictor for fitness and antibiotic sensitivity. *bioRxiv* **38**, (2019).
132. Paulsen, I. T., Cain, A. K. & Hassan, K. A. Physical enrichment of transposon mutants from saturation mutant libraries using the TraDISort approach. *Mob. Genet. Elements* **7**, 1–7 (2017).
133. O'Connor, T. J., Boyd, D., Dorer, M. S. & Isberg, R. R. Aggravating Genetic Interactions Allow a Solution to Redundancy in a Bacterial Pathogen. *Science* (80-.). **338**, 1440 LP – 1444 (2012).
134. Lee, S. Y. *et al.* Dense Transposon Integration Reveals Essential Cleavage and Polyadenylation Factors Promote Heterochromatin Formation. *Cell Rep.* **30**, 2686-2698.e8 (2020).
135. Michel, A. H. *et al.* Functional mapping of yeast genomes by saturated transposition. *Elife* **6**, e23570 (2017).
136. Guo, Y. *et al.* Integration profiling of gene function with dense maps of transposon integration. *Genetics* **195**, 599–609 (2013).
137. Zhang, C., Phillips, A. P. R., Wipfler, R. L., Olsen, G. J. & Whitaker, R. J. The essential genome of the crenarchaeal model *Sulfolobus islandicus*. *Nat. Commun.* **9**, 1–11 (2018).
138. Sarmiento, F., Mrázek, J. & Whitman, W. B. Genome-scale analysis of gene function in the hydrogenotrophic methanogenic archaeon *Methanococcus maripaludis*. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 4726–4731 (2013).
139. Carette, J. E. *et al.* Global gene disruption in human cells to assign genes to phenotypes by deep sequencing. *Nat. Biotechnol.* **29**, 542–546 (2011).
140. Carette, J. E. *et al.* Haploid Genetic Screens in Human Cells Identify Host Factors Used by Pathogens. *Science* (80-.). **326**, 1231 LP – 1235 (2009).
141. Friedrich, M. J. *et al.* Genome-wide transposon screening and quantitative insertion site sequencing for cancer gene discovery in mice. *Nat. Protoc.* **12**, 289–309 (2017).
142. Rad, R. *et al.* A conditional piggyBac transposition system for genetic screening in mice identifies oncogenic networks in pancreatic cancer. *Nat. Genet.* **47**, 47–56

- (2015).
143. Zhang, M. *et al.* Uncovering the essential genes of the human malaria parasite *Plasmodium falciparum* by saturation mutagenesis. *Science* (80-.). **360**, (2018).
 144. Smith, A. M. *et al.* Quantitative phenotyping via deep barcode sequencing. *Genome Res.* **19**, 1836–1842 (2009).
 145. Liu, X. *et al.* High-throughput CRISPRi phenotyping identifies new essential genes in *Streptococcus pneumoniae*. *Mol. Syst. Biol.* **13**, 931 (2017).
 146. Peters, J. M. *et al.* A comprehensive, CRISPR-based functional analysis of essential genes in bacteria. *Cell* **165**, 1493–1506 (2016).
 147. Lee, H. H. *et al.* Functional genomics of the rapidly replicating bacterium *Vibrio natriegens* by CRISPRi. *Nat. Microbiol.* **4**, 1105–1113 (2019).
 148. Steen, A. D. *et al.* High proportions of bacteria and archaea across most biomes remain uncultured. *ISME J.* **13**, 3126–3130 (2019).
 149. Zhu, Z., Surujon, D., Pavao, A., Bento, J. & van Opijnen, T. Forecasting bacterial survival-success and adaptive evolution through multi-omics stress- 2 response mapping, network analyses and machine learning. *bioRxiv* (2018).
doi:doi.org/10.1101/387910

Acknowledgements

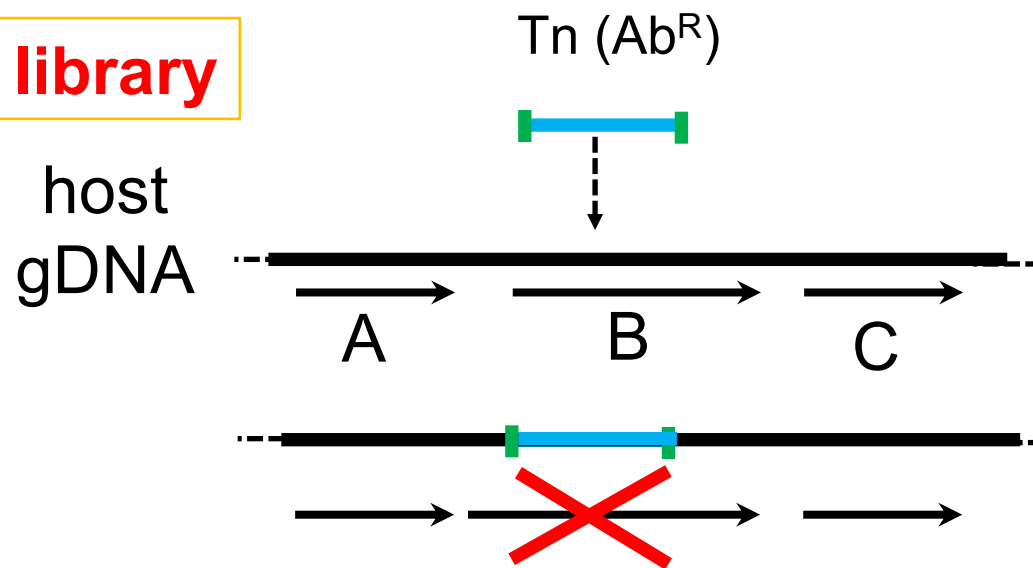
The authors would like to acknowledge funding from various sources that supported the composition of this article: ARC DECRA fellowship DE180100929 and NHMRC grant APP1159752 to A.K.C; NIH grants U01AI124302, R01AI110724 and R21AI117247 to T.O.; Bavarian antimicrobial resistance research network bayresq.net grant to L.B.; MRC grant G1100100/1 to J.P; NHMRC grants GNT1060895 and GNT1120298 to I.T.P; and NIH grant R35GM118159 to A.L.G.

Competing interests

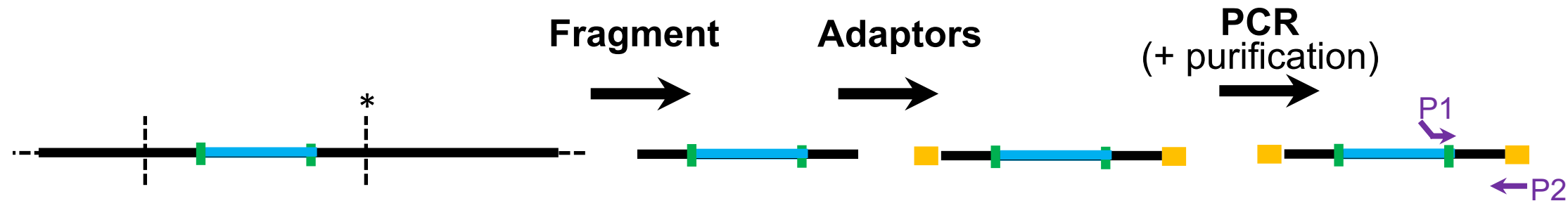
The authors declare no competing interests.

A) Create TIS library

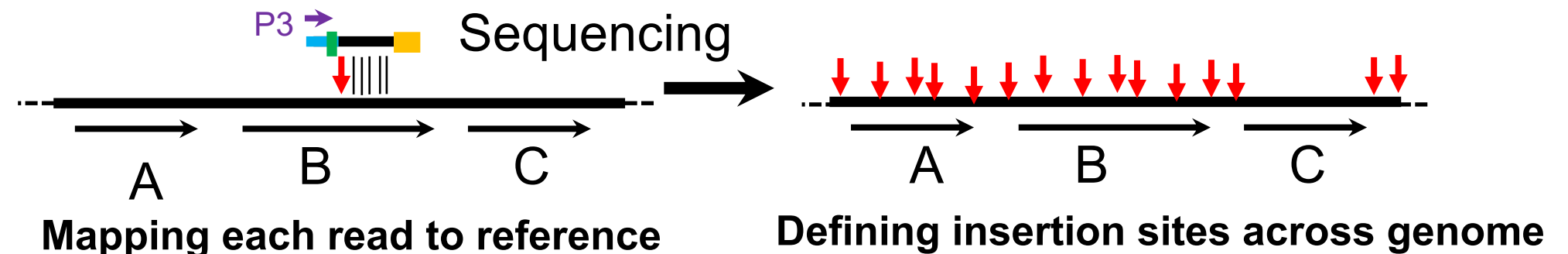
1. Create random mutants



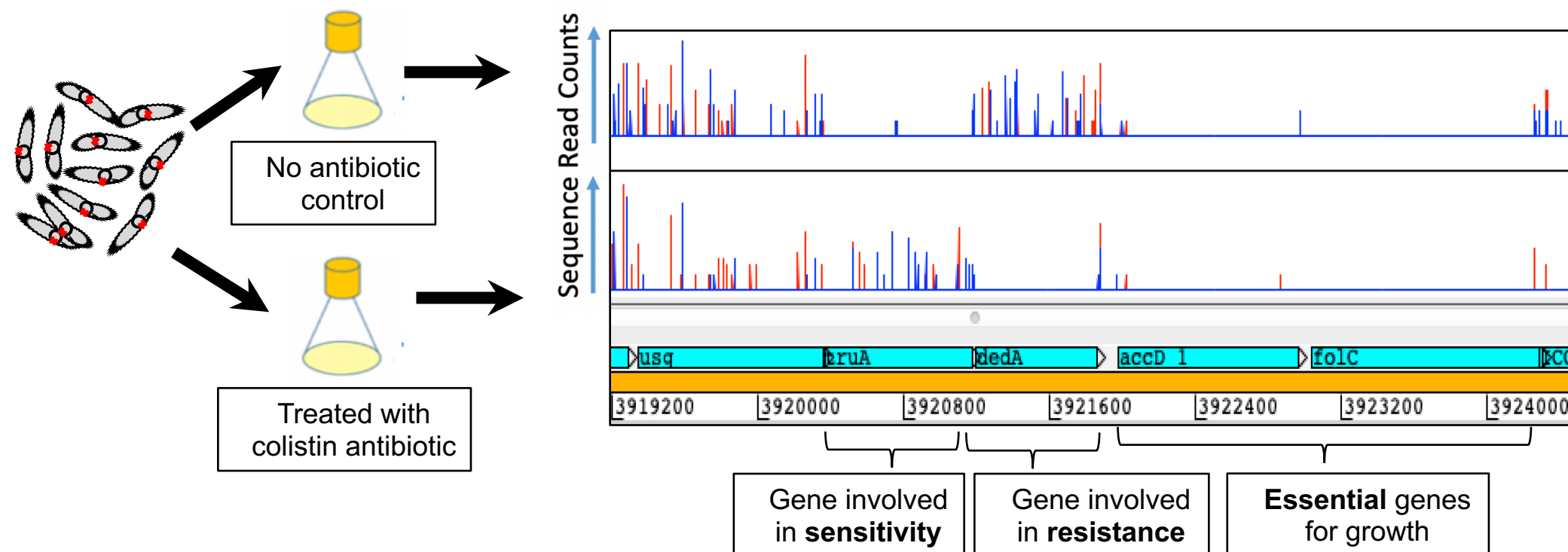
3. Fragment, adaptors and PCR amplify



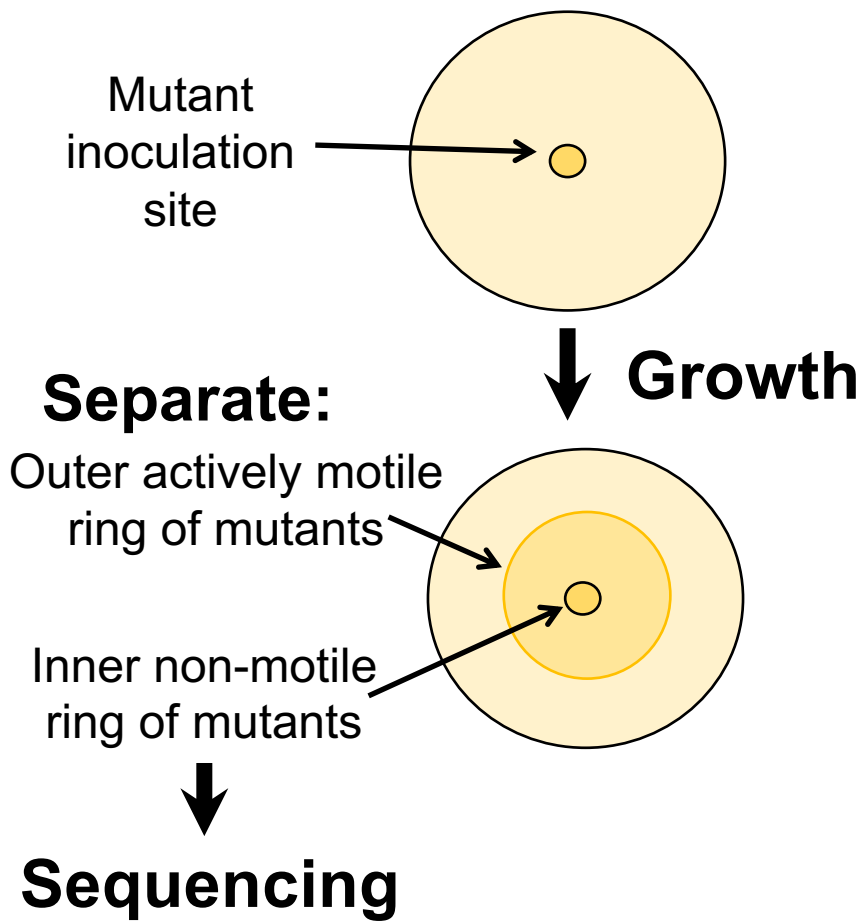
4. Sequencing and mapping



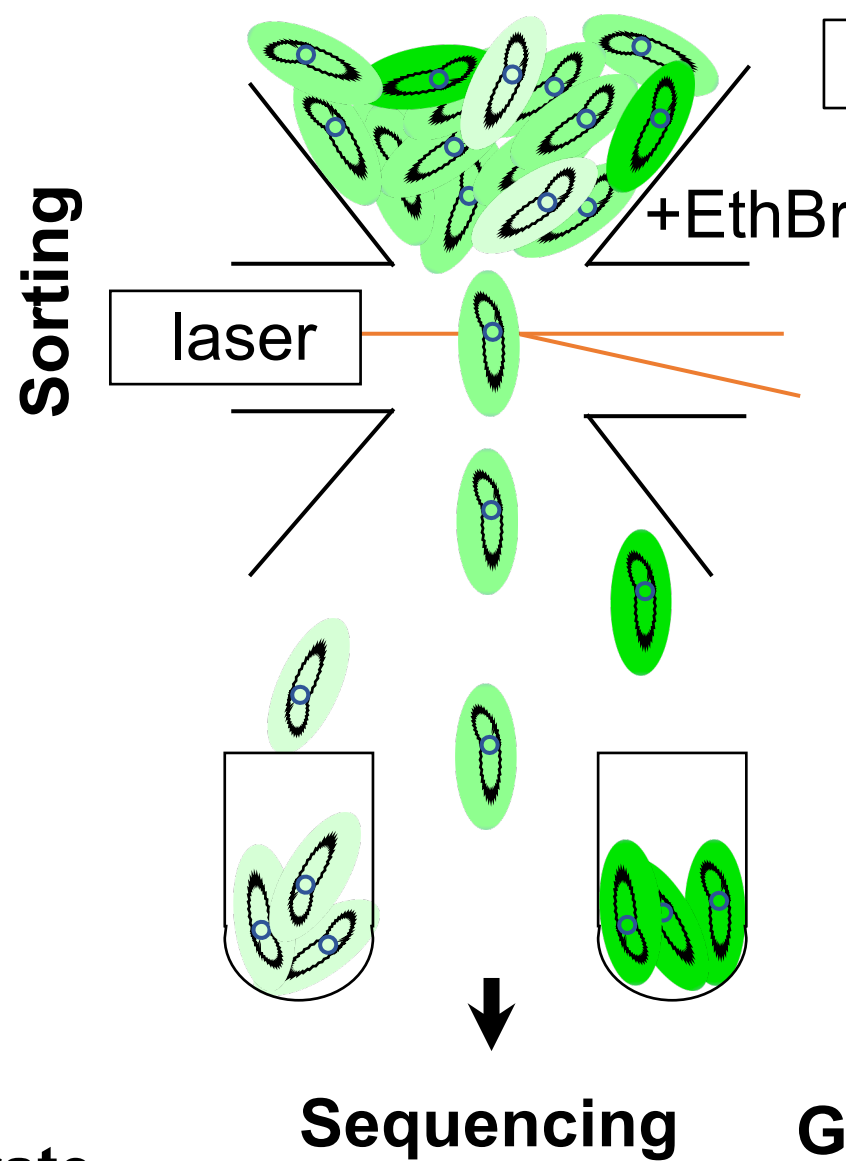
B) Challenge with condition



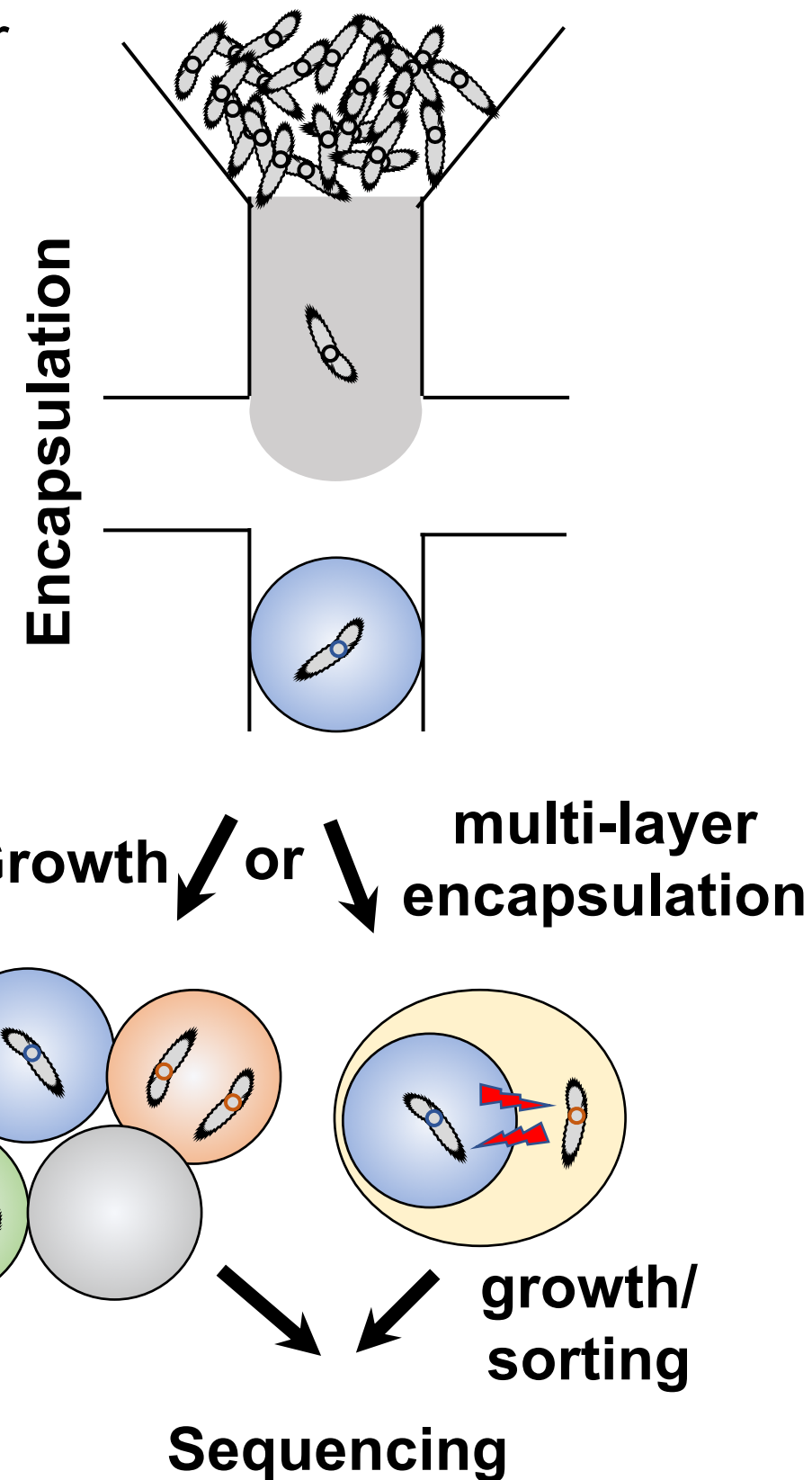
A)

Motility TIS

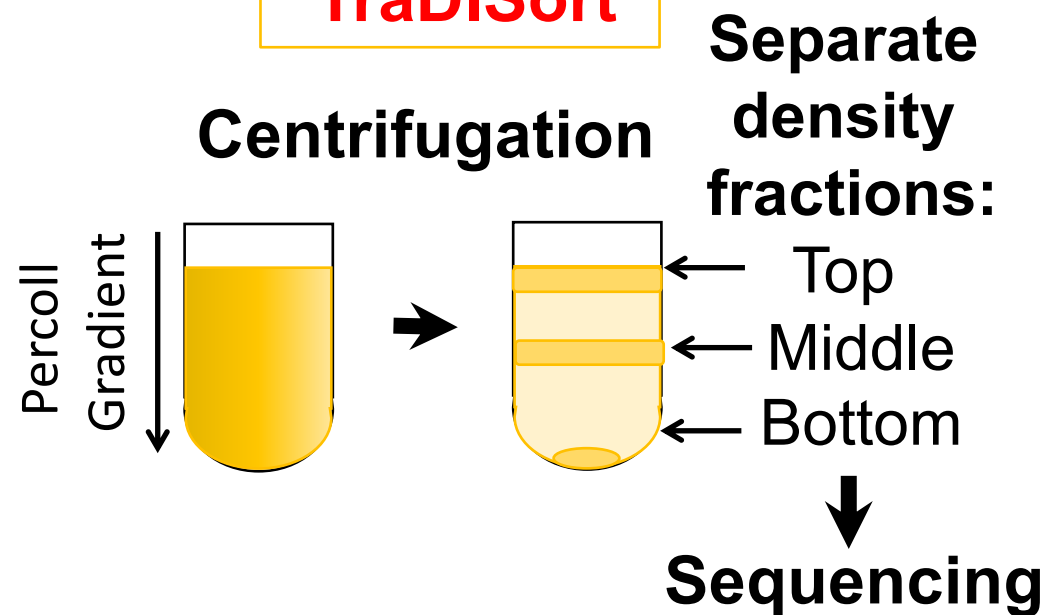
C)

TraDISort

D)

Droplet Tn-Seq

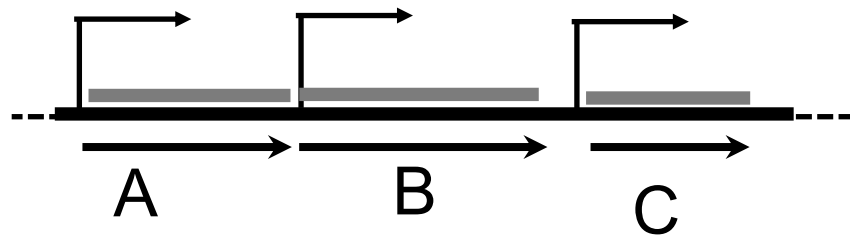
B)

Density-TraDISort

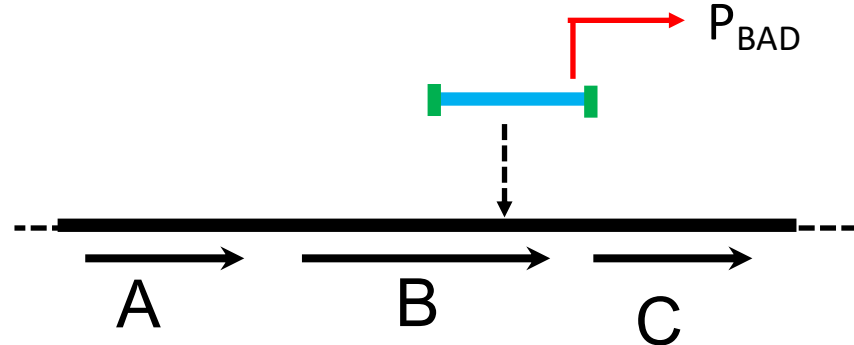
A)

Tn + outward inducible promoter

1. Wild type expression

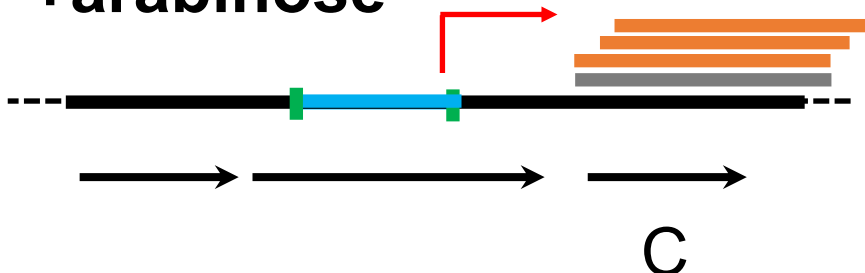


2. Tn insertion (Promoter)



3. Tn expression induction

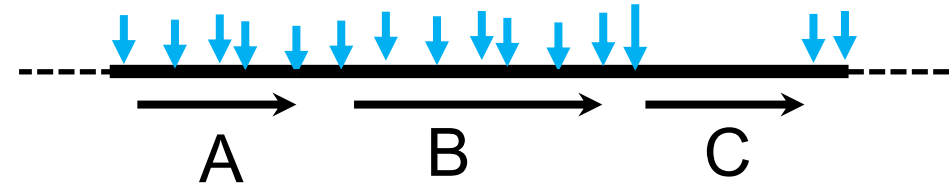
+arabinose



B)

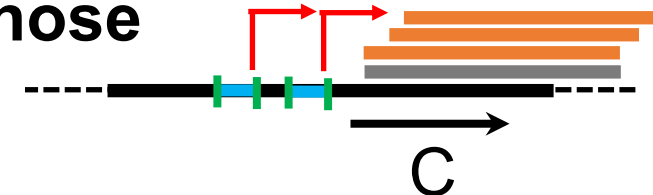
Gain of function screens

1. Tn insertions show essential genes



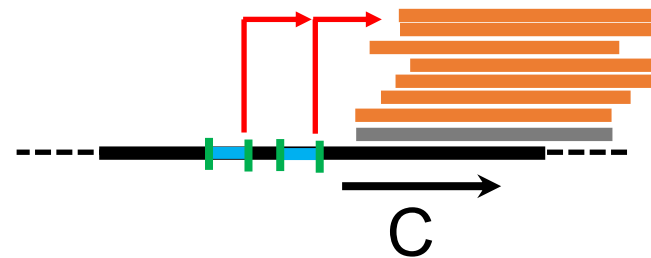
2. Overexpression of essential genes

+arabinose

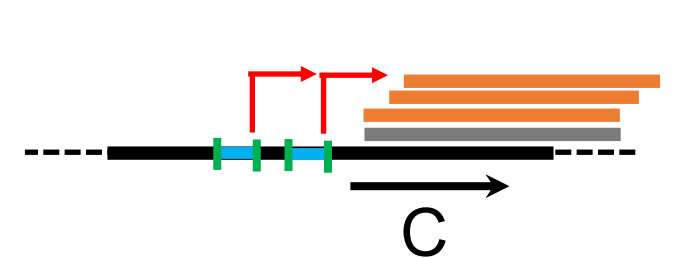


3. If essential gene C increases fitness in selection

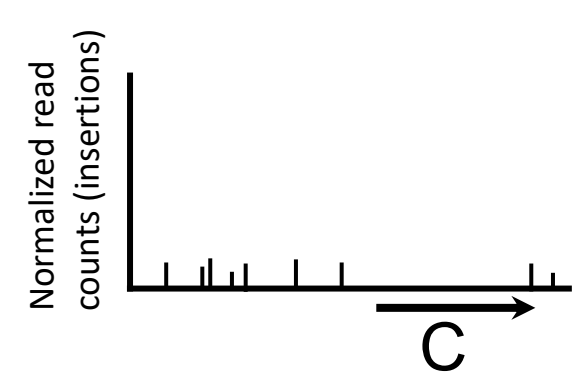
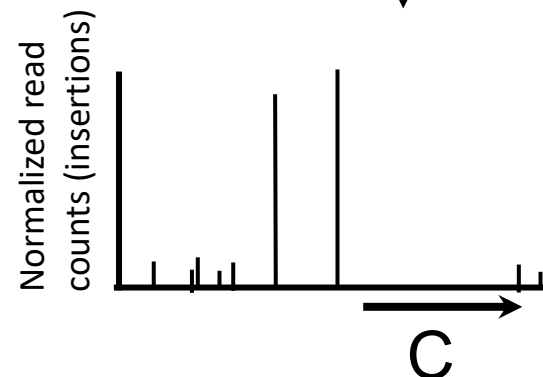
+ Selection



- Selection



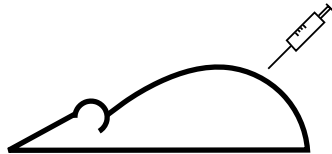
Sequencing out of Tn



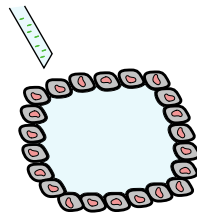
Infection Model



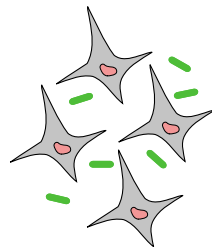
Oral



Intraperitoneal

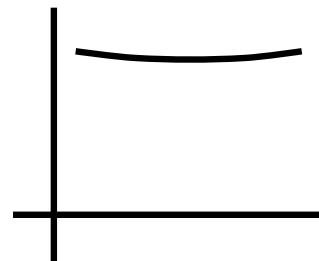
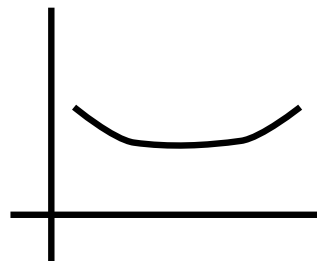
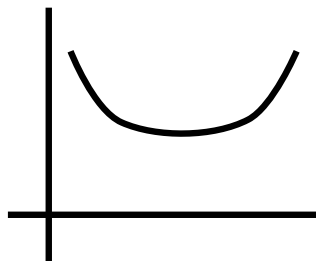
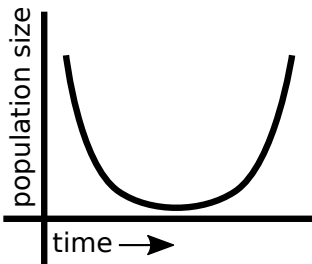


Organoid



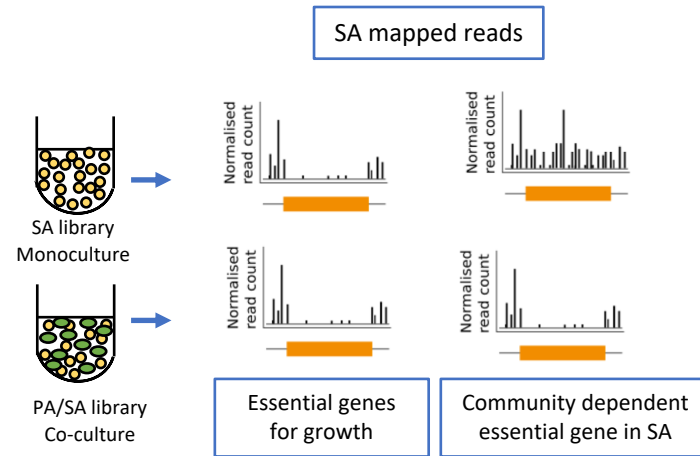
Cell Culture

Bottleneck



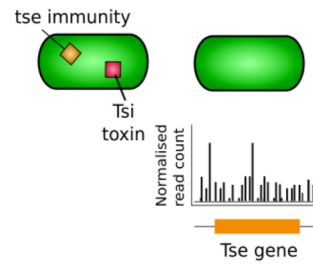
← Model Realism

A)

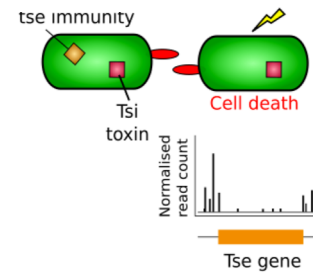


B)

PA_*H1_retS* Tn Library (Inactive T6SS)



PA_*retS* Tn Library (Active T6SS)



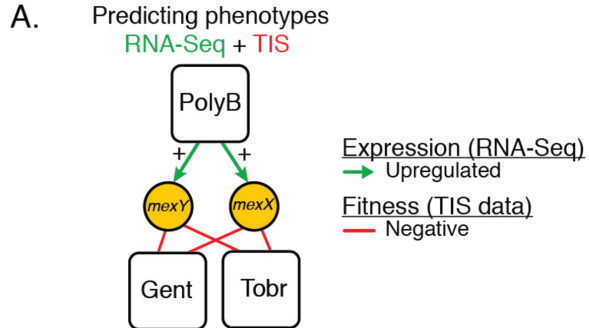
The diagram illustrates a network structure with three primary hubs: PeniG, Dapto, and Vanco. PeniG is positioned at the top, Dapto at the bottom left, and Vanco at the bottom right. These hubs are connected to a variety of peripheral nodes, represented by colored circles. PeniG connects to a cluster of orange and blue nodes at the top, a group of red, orange, and blue nodes on the left, a central cluster of blue nodes, and a group of red, orange, and green nodes on the right. Dapto connects to a group of red, orange, and blue nodes on the left, the central cluster of blue nodes, and a group of orange and blue nodes at the bottom. Vanco connects to a group of red, orange, and green nodes on the right, the central cluster of blue nodes, and a group of orange and blue nodes at the bottom. A node labeled 'Q.' is located at the bottom center, connected to the central cluster of blue nodes and Vanco. The nodes are color-coded: red, orange, blue, purple, green, yellow, and white.

The diagram illustrates a network structure. A central white node labeled **Q.** is connected to seven peripheral nodes. The connections are as follows:

- A solid black line connects **Q.** to a white rectangular node labeled **lung healthy**.
- A solid black line connects **Q.** to a red circular node.
- A dashed black line connects **Q.** to a white rectangular node labeled **lung neut-**.
- Solid black lines connect **Q.** to five blue circular nodes.
- A solid black line connects **Q.** to an orange circular node.

- regulator
- peptidoglycan synthesis
- metabolism
- cell division
- cell wall integrity
- unknown

— Negative
— Positive
--- no effect

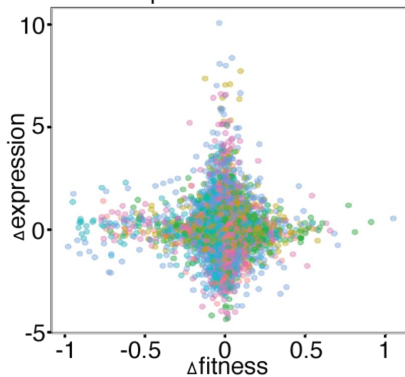


B.

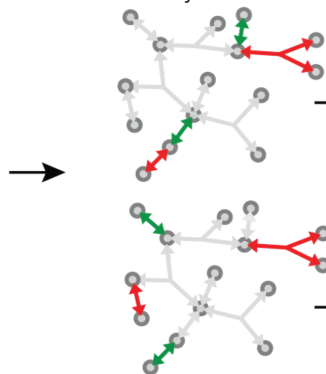
1. Measuring genomic responses



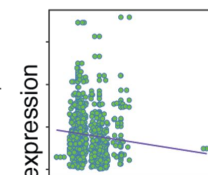
2. Expression vs. Fitness



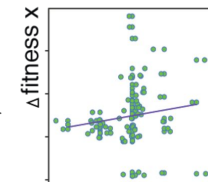
3. Data overlay on known network



4. Network relationship fitness & expression



relatively small =
coordinated/adapted



relatively large =
uncoordinated/unadapted

distance to neighbor