Published as: *Curr Protoc Microbiol.* 2010 November ; 0 1: .

Genome-Wide Fitness and Genetic Interactions Determined by Tn-seq, a High-Throughput Massively Parallel Sequencing Method for Microorganisms

Tim van Opijnen¹ and Andrew Camilli¹

¹Tufts University School of Medicine, Boston, Massachusetts

Abstract

The lagging annotation of bacterial genomes and the inherent genetic complexity of many phenotypes is hindering the discovery of new drug targets and the development of new antimicrobials and vaccines. Here we present the method Tn-seq, with which it has become possible to quantitatively determine fitness for most genes in a microorganism and to screen for quantitative genetic interactions on a genome-wide scale and in a high-throughput fashion. Tn-seq can thus direct studies in the annotation of genes and untangle complex phenotypes. The method is based on the construction of a saturated Mariner transposon insertion library. After library selection, changes in frequency of each insertion mutant are determined by sequencing of the flanking regions en masse. These changes are used to calculate each mutant's fitness. The method has been developed for the Gram-positive bacterium *Streptococcus pneumoniae*, a causative agent of pneumonia and meningitis; however, due to the wide activity of the Mariner transposon, Tn-seq can be applied to many different microbial species.

Keywords

Tn-seq; *Streptococcus pneumoniae*; genome-wide fitness; genetic interactions; transposon mutagenesis; massively parallel sequencing

INTRODUCTION

The Gram-positive bacterium *Streptococcus pneumoniae* is a common occupant of the nasopharynx and the causative agent of a variety of diseases ranging from otitis media to pneumonia, bacteremia, and meningitis. Each year, 1.5 million people succumb to infection with *S. pneumoniae* (Tuomanen et al., 2004). Due to increasing antibiotic resistance across many bacterial species and because the drug pipeline is running on empty, we are rapidly closing in on a pre-antibiotics era in terms of treatment options (Taubes, 2008). Vaccines are a promising alternative for the future to protect against bacterial infections. Against *S. pneumoniae*, there are currently two different polysaccharide-based vaccines available that protect ~60% of vaccinees. However, they are expensive, and the polysaccharide approach is limiting, as *S. pneumoniae* serotypes against which the vaccines are directed can be replaced by other serotypes that are not covered by the vaccine.

For *S. pneumoniae*, as well as for many other pathogens, it is key to gain a better understanding of the link between genotype and phenotype not only in order to be able to understand processes such as colonization, persistence, and virulence, but also to be able to identify new drug or vaccine antigen targets. One of the problems, however, is the lagging annotation of bacterial genomes. For instance, more than 30% of the genes for *S. pneumoniae* are of unknown function, and it is reasonable to assume that most genes to which functions have been ascribed are only partially understood. In addition,

microorganisms are complex systems, and a phenotype does not simply arise as the product of the individual genetic components; rather, a phenotype is a complex trait that results from these components interacting.

To aid in this quest, we have developed a widely applicable high-throughput tool for gene disruption, called Tn-seq, which has great potential for elucidating gene function, linking genotypes to phenotypes, and resolving complex pathways in microorganisms. Tn-seq is a robust and sensitive method that enables the determination of each gene's contribution to fitness in a specific environment, through massive parallel sequencing. The approach is based on the assembly of a saturated transposon insertion library. Upon growth of the library under a test condition of interest (in vitro or in vivo), insertion mutants with a lower fitness decrease in frequency in the population, while other mutants, depending on the effect of the transposon insertion, remain the same or increase in frequency. Changes in frequency are determined by sequencing the transposon flanking regions en masse on an Illumina Genome Analyzer II. Subsequently, fitness of every individual gene knockout in a bacterial genome can be calculated. Due to the quantitative nature of massive parallel sequencing, fitness of a gene knockout can be quantitatively determined and is a direct measure of the growth rate of the bacterial clone containing it. For instance, a mutant with a fitness of 0.5 translates to a two-fold higher doubling time than the wild-type (van Opijnen et al., 2009).

By applying Tn-seq to the Gram-positive bacterium *S. pneumoniae*, a causative agent of pneumonia and meningitis, we have shown that besides determining a gene's contribution to fitness, it is also possible to identify genes essential for growth and genetic interactions. Importantly, Tn-seq is not limited to *S. pneumoniae* or Gram-positive bacteria, but, due to the activity of the Mariner transposon across many different species, the method is applicable to a large number of microorganisms.

Basic Protocol 1 describes the construction of a transposon library in *S. pneumoniae*. Basic Protocol 2 provides methods for making *S. pneumoniae* starter cultures, and then creating competent cells and transforming them with linear fragments of genomic DNA from the transposon library. Basic Protocol 3 describes creation of the transposon library starter cultures and their use in selection to determine each gene's fitness in a specific environment. It also addresses sample preparation for Illumina sequencing, which is used in the determination of fitness.

CAUTION: *S. pneumoniae* is a Biosafety Level 2 (BSL-2) pathogen. Follow all appropriate guidelines and regulations for the use and handling of pathogenic microorganisms. See [UNIT 1A](#) and other pertinent resources ([APPENDIX 1B](#)) for more information. *S. pneumoniae* is a human pathogen, capable of causing a variety of diseases ranging from otitis media to pneumonia, bacteremia, and meningitis.

NOTE: *S. pneumoniae* is grown on blood agar plates or statically in rich liquid medium such as Todd Hewitt Broth supplemented with 0.5% yeast extract (THY) or Brain Heart Infusion (BHI). Liquid cultures and agar plates are incubated overnight in an incubator at 37°C with a 5% CO₂ atmosphere or in a candle jar.

TRANSPOSON LIBRARY CONSTRUCTION

The protocol describes the construction of a transposon library in *S. pneumoniae*. Through in vitro transposition mediated by the transposase MarC9, the mini-transposon *magellan6* (which is a derivative of the Himar1 Mariner transposon), encoding spectinomycin resistance, is introduced into bacterial genomic DNA linear fragments. *Magellan6* is provided on a plasmid in the reaction, and transposes by a “cut-and-paste” mechanism into the genomic DNA. The transposition event leaves single-strand gaps at each transposon-

chromosome junction, which are subsequently repaired with T4 DNA polymerase and *E. coli* DNA ligase.

Materials

S. pneumoniae genomic DNA (obtained via Qiagen Blood & Tissue Kit)

Magellan6 plasmid DNA (available from the authors on request;
andrew.camilli@tufts.edu)

2× Buffer A (see recipe)

3 M sodium acetate, pH 5.2 (*APPENDIX 2A*)

70% and 100% ethanol

MarC9: mariner transposase (see Lampe et al., 1996; add in 0.5 µl volume with concentration depending on batch; appropriate concentration should be experimentally assessed; plasmid may be obtained from the authors on request;
andrew.camilli@tufts.edu)

10× Buffer B (see recipe)

10× (10 mg/ml) bovine serum albumin (BSA; New England Biolabs)

2 mM dNTPs (*APPENDIX 2A*)

3 U/µl T4 DNA polymerase (New England Biolabs)

2.6 mM nicotine adenine dinucleotide (NAD)

10 U/µl *E. coli* DNA ligase (New England Biolabs)

30° and 75°C heating blocks

12°C and 16°C cooling blocks

Additional reagents and equipment for measuring DNA concentration (Gallagher and Desjardins, 2006)

Carry out transposition

1. Combine 1 µg *S. pneumoniae* genomic DNA, 1 µg *magellan6* (plasmid), 10 µl 2× Buffer A, and sufficient distilled water for an end volume of 19.5 µl.
2. Add 0.5 µl of MarC9 at the appropriate dilution (empirically determined).
Make sure MarC9 is stored at −80°C and that the transposase is divided into small aliquots, as thawing more than twice drastically lowers transposition efficiency.
3. Incubate at 30°C for 1 hr in a heating block.
4. Heat inactivate at 75°C for 10 min and transfer to ice for 5 min in a heating block.
5. Precipitate DNA by adding 1/10 volume of 3 M sodium acetate, pH 5.2, and 2 volumes of 100% ethanol at 4°C, incubating 20 min at 4°C, microcentrifuging at maximum speed, 4°C, to form a pellet, and washing the pellet with room temperature 70% ethanol. Air dry pellet and dissolve in 12.5 µl water.
6. Measure DNA concentration (Gallagher and Desjardins, 2006) using 1.5 µl of the DNA solution.

A yield in the range of 50 to 150 ng/µl should be obtained.

Repair transposon junctions

7. Mix the following:
 - 2 μ l 10 Buffer B
 - 2 μ l 10 $\times$ BSA
 - 1 μ l 2 mM dNTPs
 - 3 μ l H₂O
 - 1 μ l 3 U/ μ l T4 DNA polymerase
 - 11 μ l DNA from step 5.
8. Incubate at 12°C for 20 min, then inactivate the reaction at 75°C for 10 min.
9. Add 0.2 μ l 2.6 mM NAD and 1 μ l 10 U/ μ l *E. coli* DNA ligase and incubate at 16°C overnight.

If a very small-volume pipettor such as a Rainin P2 is not available, 1 μ l of 0.52 mM NAD can be used; however, the final volume of the ligation reaction will be slightly altered.

STARTER CULTURES AND TRANSFORMATION OF *S. PNEUMONIAE*

This protocol describes how to make starter cultures of *S. pneumoniae* and how these are used to make competent cells. Competent cells are transformed with linear fragments of genomic DNA containing the *magellan6* insertions to yield a transposon insertion library. Upon transformation, the *magellan6* insertions are integrated into the genome by double cross-over homologous recombination. Transformants are selected on medium containing antibiotic to which the transposon encodes resistance. Because in vitro transposition and subsequent transformation of *S. pneumoniae* are both low-frequency events, the overwhelming majority of transformants harbor single transposon insertions. The resulting library can be used in any environment of interest (in vivo or in vitro) to gain insight into each gene's contribution to fitness (survival and/or growth) in that specific environment.

Though this protocol takes advantage of the natural competence property of *S. pneumoniae*, non-competent microbial species for which in vivo transposition methods are available may also be subjected to Tn-seq. In such cases, the in vivo transposition methods would have to be adapted to use the *magellan6* mini-transposon described here. Although other transposons (e.g., Tn5) or insertion elements (e.g., suicide vector) can be used with Tn-seq, the protocol described below is specific to *magellan6*. Other transposons or insertional elements would require modifications to the procedure used to purify the transposon-chromosome junctions for sequencing.

Of significant importance in the protocol described here is the elimination of the step of plating transformed bacterial cells in soft agar layers, which has been the standard method for selecting *S. pneumoniae* transformants. Instead, transformed cells are outgrown for a short period of time to allow expression of the transposon drug marker, and then spread on the surface of antibiotic-containing blood agar plates. This change in procedure results in a dramatic increase in the dynamic range of the experiments, i.e., bacteria containing transposon insertions that have a fitness ~ 0.05 are recovered. This is in stark contrast to the use of the original soft agar "layering" method where insertion mutants are recovered that have a fitness ~ 0.5 (van Opijnen et al., 2009).

Materials

S. pneumoniae strain of choice (e.g., D39 or TIGR4)
Sheep's blood agar plates (with and without 200 µg/ml spectinomycin; see recipe)
Todd Hewitt yeast broth (THY; pH 7.3): 30 g TH base (Becton Dickinson, cat. no. 249240) per liter
Oxyrase (Oxyrase, cat. no. OB-0010)
THY (see above) containing 12% (v/v) glycerol
1 N HCl
20% (w/v) glycine (in distilled H₂O)
1 N NaOH
8% (w/v) bovine serum albumin (BSA)
1 M CaCl₂
350 ng/µl competence stimulating peptide (CSP) appropriate to the *S. pneumoniae* strain used (e.g., CSP1 for D39, CSP2 for TIGR4) (AnaSpec)
Transposon library DNA (Basic Protocol 1)
37°C, 5% CO₂ incubator
Spectrophotometer
Large (16-mm diameter) and small (13-mm diameter) glass culture tubes
Centrifuge
Heat block at 37°C

Create starter cultures for transformation

1. Streak an *S. pneumoniae* strain on a sheep's blood agar plate (without spectinomycin) and incubate overnight at 37°C with 5% CO₂.
2. Pick a single colony and grow in 250 µl THY and 5 µl Oxyrase in a 1.5 ml microcentrifuge tube for 3 hr.
3. Plate 100 µl on a blood agar plate (without spectinomycin) and incubate overnight at 37°C with 5% CO₂.
4. On the next day, resuspend all colonies from plate's surface in 3 ml THY, transfer to a large culture tube, and make volume up to 10 ml THY containing 5 µl/ml Oxyrase. Incubate for 2 to 3 hr at 37°C with 5% CO₂ until the OD at 600 nm is ~0.3.
5. Centrifuge 8 min at 3000 × g, room temperature.
6. Discard the supernatant and resuspend the cell pellet in 5 ml THY containing 12% glycerol.
7. Aliquot into twenty 1.5-ml microcentrifuge tubes and store the starter cultures up to 9 months at -80°C.

Carry out transformation

8. Thaw a starter culture and transfer to 7.5 ml THY in a large glass culture tube (do not add antibiotics or Oxyrase).
9. Add 97.5 μ l 1 N HCl and 18.8 μ l 20% glycine. Allow to grow for 1.5 to 2 hr and measure the OD at 600 nm (should be between 0.03 and 0.1).
10. Transfer a volume of the culture into a small culture tube and add prewarmed THY to final volume of 1 ml so that the OD is adjusted to ~0.03 (e.g., if OD of the original culture is 0.06 then add 500 μ l THY to 500 μ l culture).
11. Place culture tube in heating block at 37°C.
12. Add the following in order, briefly and gently vortexing after each addition:
 - 10 μ l 1 N NaOH
 - 25 μ l 8% BSA
 - 1 μ l 1 M CaCl.Make sure to add these components separately. They will precipitate out if mixed beforehand.
13. Add 1.5 μ l CSP, immediately start the timer, and place the culture tube back into 37°C incubator with 5% CO₂.
14. Precisely 14 min after adding CSP2 or 15 min after adding CSP1, add 22 μ l DNA from Basic Protocol 1, step 9.
 - Note that the type of CSP and incubation time depends on the *S. pneumoniae* strain used. For instance use CSP1 and 15 min of incubation for strain D39, and use CSP2 and 14 min of incubation for TIGR4 (Pozzi et al., 1996; Hsieh et al., 2006).
15. Place back in 5% CO₂ incubator at 37°C for 45 min.

Prepare transposon library starter cultures

16. Transfer 1 ml of transformed culture into a microcentrifuge tube and centrifuge 8 min at 3000 $\times$ g, room temperature.
17. Discard 900 μ l of supernatant, resuspend the bacterial cell pellet in the leftover 100 μ l of THY, and plate on a blood agar plate containing 200 μ g/ml spectinomycin.
 - Note that a cell pellet is often not visible.
18. Incubate overnight at 37°C with 5% CO₂.
19. Resuspend colonies in 3 ml THY, transfer to a culture tube, and bring to 10 ml with THY containing 200 μ g/ml spectinomycin and 5 μ l/ml Oxyrase.
20. Incubate for 30 min to 1 hr to an OD at 600 nm of 0.3.
21. Spin down cells by centrifuging 8 min at 3000 $\times$ g, room temperature, discard supernatant, and dissolve pellet in 5 ml THY containing 12% glycerol.
22. Aliquot into 20 microcentrifuge tubes and store the transposon library starter cultures at -80°C.

TRANSPOSON LIBRARY SELECTION AND ILLUMINA SEQUENCING

SAMPLE PREPARATION

This protocol describes how a transposon library starter culture can be used to determine each gene's fitness in a specific environment. In short, a starter culture is thawed and briefly cultured (30 min), a sample is taken for DNA isolation (t_1), and a part of the culture is used to seed an environment of interest. This latter culture is incubated for several hours during which the library of strains is selected upon (e.g., for growth in a liquid medium), then another sample is taken for DNA isolation (t_2). Next, the DNA from both time points is prepared for Illumina sequencing. Here we use THY as an example of a selective environment. In this example, we are thus determining which genes are important and which are dispensable for growth in THY. However, this environment can be changed to any environment of interest including infection in a mouse model.

A key step in this protocol is digestion of the DNA with the type IIS restriction enzyme *MmeI*. The inverted repeat sequences present at the left and right ends of *magellan6* each contain an *MmeI* recognition sequence. *MmeI* cuts 20 bp downstream of its recognition site, leaving a random two base 3' overhang. *MmeI* thus generates DNA fragments, of which a subset contain the *magellan6* inverted repeat and ~20 bp of flanking DNA. By subsequently ligating on an adapter with a random two base 5' overhang, DNA fragments with the same spacing between adapter and *magellan6* inverted repeat are created. These fragments can be amplified in a PCR reaction in an unbiased manner and then sequenced en masse.

Materials

Transposon library starter culture (Basic Protocol 2)

Todd Hewitt yeast broth (THY; pH 7.3): 30 g TH base (Becton Dickinson, cat. no. 249240) per liter

Spectinomycin (Sigma-Aldrich)

Oxyrase (Oxyrase, cat. no. OB-0010)

Qiagen Blood & Tissue Kit

Adapter oligonucleotides (Table 1E.3.1)

1 mM Tris·Cl, pH 8.3 (APPENDIX 2A)

MmeI restriction enzyme (New England Biolabs, cat. no. R0637S)

32 mM S-adenosyl methionine (SAM; New England Biolabs)

Buffer 4 (New England Biolabs)

Calf intestinal phosphatase (CIP; New England Biolabs, cat. no. M0290S)

24:23:1 (v/v/v) phenol:chloroform:isoamyl alcohol

100% and 70% ethanol

400 U/μl T4 DNA ligase (New England Biolabs) and T4 DNA ligase buffer

25 nM dNTP mix

PFU Ultra polymerase and PFU Ultra buffer (Stratagene)

10 μM PCR primers (Table 1E.3.2)

SeaKem LE agarose

50-bp DNA ladder
Large (16 mm diameter) glass culture tubes
Spectrophotometer
37°C, 5% CO₂ incubator
Heat block
16°C cooling block
PCR thermal cycler

Additional reagents and equipment for isolation of genomic DNA from bacteria (Qiagen Blood & Tissue Kit; also see Wilson, 1997), phenol/chloroform/isoamyl alcohol precipitation of DNA (Moore and Dowhan, 2002), and agarose gel electrophoresis (Voytas, 2000)

Select library

1. Thaw a transposon library starter culture and transfer to 4 ml THY (add spectinomycin to 200 µg/ml, and 20 µl Oxyrase). Allow to grow for 30 to 60 min and measure OD at 600 nm, which should be ~ 0.1.
2. Seed 10 ml THY at an OD at 600 nm of 0.005 and incubate at 37°C with 5% CO₂ (the selection). In addition, take a small sample to enumerate cfu in the culture at t_1 .
3. Spin down the remaining culture that was used to seed the selective culture 8 min at 3000 × *g* room temperature, and discard supernatant. Isolate genomic DNA using the Qiagen Blood & Tissue Kit (also see Wilson, 1997) immediately, or store cell pellet at -20°C for later DNA isolation.

This represents the t_1 DNA.

4. When the selective culture reaches an OD at 600 nm of approximately 0.6, take a small sample to enumerate cfu, spin down the rest of the culture 8 min at 3000 × *g*, room temperature, discard supernatant, and isolate genomic DNA using the Qiagen Blood & Tissue Kit (also see Wilson, 1997).

This represents the t_2 DNA.

Prepare sample for sequencing

5. Prepare the adapter by bringing both oligonucleotides (Table 1E.3.1) to 0.2 mM with 1 mM Tris-Cl, pH 8.3. Mix equal volumes of each in a microcentrifuge tube and place in a heat block at 96°C for 2 min.

See Figure 1E.3.1 and Tables 1E.3.1 and 1E.3.2 for details on how the adapter and primers combine and result in DNA molecules that can be sequenced on an Illumina GAII. Note that the two oligonucleotides that form the adapter contain a four-nucleotide barcode sequence (see Table 1E.2.2). By varying this sequence, multiple samples can be mixed into a single Illumina flow cell lane, which can be separated after sequencing based on the barcode. For instance, by varying the barcode for t_1 and t_2 , these two samples can be mixed and sequenced in the same flow cell lane.

6. To anneal the two oligonucleotides, place the part of the heat block that holds the microcentrifuge tubes on the bench and let it cool to room temperature (~10 to 20 min).

7. Store the adapter at -20°C until use.
8. Digest genomic DNA from t_1 and t_2 with the following mixture at 37°C for 2.5 hr:
 - 3 μg *S. pneumoniae* DNA
 - 3 μl *MmeI*
 - 0.44 μl 32 mM SAM
 - 20 μl Buffer 4
 - H_2O to 200 μl .
9. Add 2 μl CIP and incubate at 37°C for 1 hr.
10. Extract the DNA with phenol:chloroform:isoamyl alcohol, precipitate with 95% ethanol, wash with 70% ethanol, and resuspend in 25 μl distilled water.

Extraction and precipitation of DNA are detailed in Moore and Dowhan (2002). Note that extraction and precipitation of the DNA greatly improves the efficiency of the next two steps.
11. Ligate adapter by mixing:
 - 25 μl DNA from step 10
 - 1 μl adapter from step 7
 - 1 μl 400 U/ μl T4 DNA ligase
 - 3 μl T4 DNA ligase buffer
 - Incubate at 16°C overnight.
12. Prepare the following PCR mix to amplify the 120-bp target:
 - 5 μl PFU Ultra buffer
 - 0.5 μl 25 mM dNTPs
 - 1 μl 10 μM primer P1_M6_MmeI
 - 1 μl 10 μM Gex PCR Primer 2
 - 0.5 μl PFU Ultra
 - 40.5 μl H_2O
 - 1.5 μl ligation mix from step 11.
13. Carry out the following PCR program:

1 cycle:	30 sec	95°C	(initial denaturation)
18-22 cycles:	10 sec	95°C	(denaturation)
	25 sec	55°C	(annealing)
	5 sec	72°C	(extension)
1 cycle:	10 min	72°C	(final extension).

The goal is to perform as few cycles as possible to remain in the linear range of the PCR.
14. Load and run all of the PCR reaction products on a 1.8% agarose gel (Voytas, 2000) together with a 50-bp DNA ladder.

15. Excise the 120-bp band, avoiding primer and primer-dimer bands.
16. Purify the DNA from the excised gel slice (e.g., using a Qiagen Gel purification kit) and elute in 50 μ l distilled water.
17. Determine yield and assess quality of the DNA on a Bioanalyzer (Agilent).
The yield should be ~ 3 ng/ μ l and the Bioanalyzer should show a distinct peak in the 120-bp size range.
18. Sequence the t_1 and t_2 samples on an Illumina Genome Analyzer II according to the manufacturer's protocol.

After sequencing, reads are mapped to the genome and the change in the number of reads per insertion over time (the change from t_1 to t_2) can be used to accurately determine fitness for each gene. A detailed description of this method and data analysis can be found elsewhere (van Opijnen et al., 2009).

REAGENTS AND SOLUTIONS

Use deionized, distilled water in all recipes and protocol steps. For common stock solutions, see **APPENDIX 2A**; for suppliers, see **SUPPLIERS APPENDIX**.

Buffer A, 2x

5 μ l 80% (v/v) glycerol
4 μ l 20 mM DTT
4 μ l 10 mg/ml BSA (New England Biolabs)
8 μ l 5 $\times$ salt (see recipe)
Prepare fresh just before use

Buffer B, 10x

500 mM Tris-Cl, pH 7.8 (*APPENDIX 2A*)
100 mM MgCl₂
10 mM DTT
Store up to 1 year at -20°C

Salt, 5x

100 μ l 125 mM HEPES, pH 7.9
11.5 μ l 5 M NaCl
5.75 μ l 1 M MgCl₂
Prepare fresh just before use

Sheep's blood agar plates

42.5 g blood agar base number 2 (Sigma, cat. no. B1676; prepare according to manufacturer's instructions)
1 liter H₂O
Autoclave

Cool to ~50°C

Add 50 ml defibrinated sheep's blood (Northeast Laboratory Services, <http://www.nelabservices.com/>)

Add 200 µg/ml spectinomycin (Sigma-Aldrich) if required in protocol step

Pour into sterile 100 × 15-mm plates

Store up to 2 months at 4°C in the dark

COMMENTARY

Background Information

Understanding how a genotype leads to a specific phenotype is key to understanding how an organism functions and survives. For some time now, it has been clear that the link between genotype and phenotype is not a linear one—rather, interactions among and between genetic components and the environment create a complex pathway leading to a specific phenotype. The result is intriguing in that through these interactions a robust organism emerges that is able to withstand many kinds of genetic and environmental disturbances. Such interactions are described in terms of epistasis, which is a phenomenon whereby the effects of a given gene on a biological trait are masked or enhanced by one or more other genes (Kitano, 2004; Moore, 2005). In the last eight years, such dependencies between genes have been extensively explored in *Saccharomyces cerevisiae* by identifying synthetic genetic interactions through systematically knocking out all possible double combinations of genes. This method has identified specific networks of genes responsible for important pathways such as DNA integrity (Ooi et al., 2003; Pan et al., 2006), mRNA export (Hieronymus et al., 2004), drug resistance (Begley et al., 2002; Giaever et al., 2004; Dudley et al., 2005; Parsons et al., 2006), and phosphorylation (Fiedler et al., 2009).

Three approaches have been developed to identify genetic interactions in *S. cerevisiae*: synthetic genetic array (SGA) technology (Tong et al., 2001), diploid-based synthetic lethality analysis on microarrays (dSLAM; Pan et al., 2004), and epistatic miniarray profile (E-MAP). Where SGA and dSLAM are able to map synthetic sick/lethal (SSL) relationships, E-MAP is more sensitive and able to measure a larger range of epistatic interactions, including positive interactions (Schuldiner et al., 2005; Collins et al., 2006, 2007; Roguev et al. 2008; Wilmes et al., 2008). These methods all make use of robotic automation to construct double mutants by mating. A query strain, with a drug-resistance marker replacing a gene of interest, is crossed to an arrayed collection of single-gene deletion strains of opposite mating type marked with a different selectable marker. After meiosis, sporulation, and selection for double mutants, genetic interactions can be identified by measuring colony size. Meiotic assortment is not applicable to bacteria. However, two analogous methods have been developed for *E. coli*: GIANT coli and eSGA, both of which make use of conjugation to drive genetic exchange of a marked query gene mutation from an *E. coli* Hfr donor strain into a genome-wide, high-density-arrayed collection of *E. coli* F-recipient single-gene mutants. Robotic pinning, dual-marker selection, and colony imaging are used to identify colony growth defects in the resulting double mutants. However, the applicability of these methods to a wide range of microorganisms is greatly limited, primarily because there are arrayed collections of single-gene knockouts for only a handful of species. A more generally applicable method to identify genetic interactions in bacteria is TraSH (Joshi et al., 2006), as well as related genomic footprinting methods (Girgis et al., 2007). TraSH is based on the construction of a transposon library, the generation of probes from outward reading T7 promoters located at each end of the transposon, and subsequent detection by microarray. However, the accuracy of this and related methods is somewhat

limited, and the use of a microarray restricts its use to the strain for which the microarray is developed.

An overview of Tn-seq—With the development of Tn-seq it has become possible to quantitatively determine fitness for most genes in a microorganism and to screen for quantitative genetic interactions on a genome-wide scale and in a high-throughput fashion. In *S. pneumoniae*, for which the method was originally developed, Tn-seq works as follows (Fig. 1E.3.2A-C). A gene disruption library is constructed by first transposing the mini-transposon *magellan6* into bacterial genomic DNA in vitro using the transposase MarC9 and subsequently transforming bacteria with the transposed DNA. The result is a pool of strains where each bacterium contains a single transposon insertion in its genome. DNA is isolated from a portion of the bacterial pool (t_1); another portion is used to seed a culture on which selection is performed (in vitro or in vivo); after selection, bacteria are recovered and DNA is isolated again (t_2). To accomplish sequencing of only the regions flanking the transposon, a single nucleotide in the transposon's inverted repeats was mutated, thereby introducing an *MmeI* restriction site. DNA from both time points is digested with *MmeI*, which cuts 20 bp downstream of its recognition site leaving a two-base overhang to which an adapter is ligated. A PCR is performed with one primer having complementarity to the adapter and the other complementary to the inverted repeat sequence. The resulting PCR product is 120 bp long, with ~20 bp of bacterial specific DNA flanked by Illumina specific sequences, which enable sequencing. After sequencing, different samples are split based on barcode sequence; reads are mapped to the genome and counted for each insertion, thus allowing fitness to be calculated.

Obtaining a robust fitness value and sequencing multiple samples in a single flow cell lane

The application of massively parallel sequencing in Tn-seq makes it possible to determine the exact location of each transposon insertion in a complex transposon insertion strain library and simultaneously follow the quantitative change during selection of each of these insertion strains in the population over time. The quantitative change is then used to determine the fitness effect of each transposon insertion. However, what makes Tn-seq so robust is that a gene's fitness is based on the average effect of multiple independent transposon insertions into a gene. In Figure 1E.3.3A, fitness effects of multiple independent transposon insertions from four independent libraries and their overall means are plotted for the region covering genes SP0456 to SP0460, illustrating how robust fitness values are obtained for individual genes.

In this light, it is important to consider the number of transposon insertions and samples that are sequenced in a single flow cell lane. When we first started to perform Tn-seq, we generally obtained 6×10^6 reads per flow cell lane and combined three libraries, containing 10,000 insertion strains each, in a single lane. This means that, on average, we obtained 200 reads per insertion, which gives good resolution to determine small changes in frequency of an insertion mutant. Now, however, the Illumina GA II runs more efficiently (partly due to experience and partly due to improvements in the platform), and we generally obtain 16×10^6 reads, which enables us to combine six independent libraries each containing 10,000 mutants. We have thus increased the cost effectiveness of Tn-seq by two-fold.

Library construction improvement—Since the original publication of Tn-seq in *Nature Methods* (van Opijnen et al., 2009), we have made a significant improvement to the transformation protocol during construction of the transposon insertion library. Originally, the transformation protocol for *S. pneumoniae* called for plating the transformants in a soft agar overlay on top of a blood agar plate, and then overlaying that with a layer containing antibiotic to select for transformants. Embedded colonies appear after overnight growth,

each of which contains a transposon insertion. Transformants then have to be separated from the soft agar, which is accomplished by chopping up the agar, diluting it in THY, and then incubating for several hours. We hypothesized that due to out-competition during these early steps of library construction, mutants with fitness values less than 0.5 were not identified, and changes to the protocol that reduce the number of generations prior to selection at t_1 would increase the dynamic range of the method. The hypothesis turned out to be correct and the solution extremely simple. We found that, after a short outgrowth period in liquid medium, the transformed cells can be directly spread onto a blood agar plate containing the selective antibiotic without loss of transformation efficiency. The next morning, antibiotic-resistant transformants can simply be washed off of the surface of these plates (see steps 16 to 19 in Basic Protocol 2).

Genetic interactions, different environments, strains, and species—In addition to its utility in determining the fitness contribution of genes, Tn-seq is well suited for subsequently identifying genome-wide genetic interactions. To identify interactions between a gene of interest (query gene) and the rest of the genome, the query gene is deleted, and subsequently a saturated transposon library is introduced in this query gene deletion background. The fitness of each double mutant is determined by Tn-seq, and compared to the expected fitness value, which is the product of the individual fitness values previously determined using Tn-seq in the wild-type background. A double mutant fitness value that deviates from the expected value reveals a genetic interaction (Avery and Wasserman, 1992; Drees et al., 2005; Collins et al., 2007; Fiedler et al., 2009; van Opijnen et al., 2009).

When performed in a systematic manner, it is possible to untangle complex phenotypes and assign genes of unknown function to molecular pathways through interaction-profile association with known genes (Tong et al., 2001; Tong, 2004; St. Onge et al., 2007; Carter et al., 2009; van Opijnen et al., 2009). Moreover, Tn-seq can be easily applied to different environments (e.g., different growth media, chemical or physical stresses, or infection models), as well as to different strains of a particular species. Finally, the Himar1 Mariner transposon can be used in many other microorganisms including *E. coli* (Lampe et al., 1999), *Staphylococcus aureus* (Bae et al., 2008), *Haemophilus influenzae* (Akerley et al., 1998), and *Mycobacterium tuberculosis* (Sasseti et al., 2003).

Critical Parameters and Troubleshooting

The most important point to consider when performing this method in *S. pneumoniae* is whether the presence of a capsule is necessary for the experiments to be successful. If the capsule is of no interest, an acapsular strain is preferred, for the simple reason that transformation efficiency is at least an order of magnitude more efficient, and therefore large transposon libraries are more easily obtained. However, by scaling up the number of transformations, we have also been successful in making large transposon libraries in capsular strains.

In Basic Protocol 2, step 8, it is very important to have fast-growing starter cultures; the faster a starter culture grows, the higher the transformation efficiency becomes. For best results, a starter culture should reach the appropriate OD within 2 hr of seeding the growth medium.

In Basic Protocol 3, make sure the genomic DNA is clean, which increases the efficiency of *MmeI* restriction digestion (step 8). In addition, it is essential to perform the phenolchloroform step of Basic Protocol 3 (step 10); omitting this step will result in failure of the following steps.

Because *S. pneumoniae* is naturally competent, it is possible to perform the transposition step in vitro, but this is not possible for many other organisms that lack efficient transformation procedures. However, in vivo transposition constructs for Mariner have been published (Akerley et al., 1998; Lampe et al., 1999; Sassetti et al., 2003; Bae et al., 2008), which makes the described method and approach applicable to those organisms.

Anticipated Results

Robust fitness effects for individual genes are obtained because they are measured by averaging across multiple independent transposon insertions in each gene (Fig. 1E.3.3A). Due to improvements made to the transformation protocol during construction of the transposon insertion library provided in this unit, the dynamic range of Tn-seq in *S. pneumoniae* is significantly increased. Figure 1E.3.3B shows a Pearson correlation of two independent libraries (10,000 insertion strains each) that were constructed with the improved transformation protocol, which illustrates that reproducibility is very high ($R^2 = 0.92$) and mutants with fitness <0.5 are observed. In addition, Figure 1E.3.3C shows the distribution of fitness for all genes in the *S. pneumoniae* TIGR4 genome based on four independent libraries (containing 10,000 insertion mutants each), which also confirms that with the improved transformation protocol mutants with fitness <0.5 are identified.

Time Considerations

The complete procedure, excluding sequencing, can be performed in 8 days: Basic Protocol 1 (transposon library construction) takes 1 day. Basic Protocol 2 takes 4 days total, including incubation times: making and freezing the starter cultures for the transformation, (2 days), carrying out the transformation (1 day), and preparing and freezing the transposon library starter cultures (1 day). Basic Protocol 3 takes 3 days: selecting the library (1 day) and preparing the library for sequencing (2 days). The sequencing can take anywhere between 2 to 4 days, depending on the Illumina GAI software versions used.

Literature Cited

- Akerley BJ, Rubin EJ, Camilli A, Lampe DJ, Robertson HM, Mekalanos JJ. Systematic identification of essential genes by in vitro Mariner mutagenesis. *Proc. Natl. Acad. Sci. U.S.A.* 1998; 95:8927–8932. [PubMed: 9671781]
- Avery L, Wasserman S. Ordering gene function: The interpretation of epistasis in regulatory hierarchies. *Trends Genet.* 1992; 8:312–316. [PubMed: 1365397]
- Bae T, Glass EM, Schneewind O, Missiakas D. Generating a collection of insertion mutations in the *Staphylococcus aureus* genome using *bursa aurealis*. *Methods Molec. Biol.* 2008; 416:103–116. [PubMed: 18392963]
- Begley TJ, Rosenbach AS, Ideker T, Samson LD. Damage recovery pathways in *Saccharomyces cerevisiae* revealed by genomic phenotyping and interactome mapping. *Molec. Cancer Res.* 2002; 1:103–112. [PubMed: 12496357]
- Carter GW, Galas DJ, Galitski T. Maximal extraction of biological information from genetic interaction data. *PLoS Comput. Biol.* 2009; 5:e1000347. [PubMed: 19343223]
- Collins S, Schuldiner M, Krogan N, Weissman J. A strategy for extracting and analyzing large-scale quantitative epistatic interaction data. *Genome Biol.* 2006; 7:R63. [PubMed: 16859555]
- Collins S, Miller K, Maas N, Roguev A, Fillingham J, Chu C, Schuldiner M, Gebbia M, Recht J, Shales M, Ding H, Xu H, Han J, Ingvarsdottir K, Cheng B, Andrews B, Boone C, Berger S, Hieter P, Zhang Z, Brown G, Ingles C, Emili A, Allis C, Toczyski D, Weissman J, Greenblatt J, Krogan N. Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature.* 2007; 446:806–810. [PubMed: 17314980]
- Drees BL, Thorsson V, Carter GW, Rives AW, Raymond MZ, Avila-Campillo I, Shannon P, Galitski T. Derivation of genetic interaction networks from quantitative phenotype data. *Genome Biol.* 2005; 6:R38. [PubMed: 15833125]

- Dudley AM, Janse DM, Tanay A, Shamir R, Church GM. A global view of pleiotropy and phenotypically derived gene function in yeast. *Mol. Syst. Biol.* 2005; 1:1–11.
- Fiedler D, Braberg H, Mehta M, Chechik G, Cagney G, Mukherjee P, Silva AC, Shales M, Collins SR, von Wageningen S, Kemmeren P, Holstege FC, Weissman JS, Keogh M, Koller D, Shokat KM, Krogan NJ. Functional organization of the *S. cerevisiae* phosphorylation network. *Cell.* 2009; 136:952–963. [PubMed: 19269370]
- Gallagher SR, Desjardins PR. Quantitation of DNA and RNA by absorption and fluorescence spectroscopy. *Curr. Protoc. Mol. Biol.* 2006; 76:A.3D.1–A.3D.21.
- Giaever G, Flaherty P, Kumm J, Proctor M, Nislow C, Jaramillo DF, Chu AM, Jordan MI, Arkin AP, Davis RW. Chemogenomic profiling: Identifying the functional interactions of small molecules in yeast. *Proc. Natl. Acad. Sci. U.S.A.* 2004; 101:793–798. [PubMed: 14718668]
- Girgis H, Liu Y, Ryu W, Tavazoie S. A comprehensive genetic characterization of bacterial motility. *PLoS Genet.* 2007; 3:e154.
- Hieronymus H, Yu M, Silver PS. Genome-wide mRNA surveillance is coupled to mRNA export. *Genes Dev.* 2004; 18:2652–2662. [PubMed: 15489286]
- Hsieh Y-C, Wang J-T, Lee W-S, Hsueh P-R, Shao P-L, Chang L-Y, Lu C-Y, Lee C-Y, Huang F-Y, Huang L-M. Serotype competence and penicillin resistance in *Streptococcus pneumoniae*. *Emerging Infect. Dis.* 2006; 12:1709–1714. [PubMed: 17283621]
- Joshi SM, Pandey AK, Capite N, Fortune SM, Rubin EJ, Sassetti CM. Characterization of mycobacterial virulence genes through genetic interaction mapping. *Proc. Natl. Acad. Sci. U.S.A.* 2006; 103:11760–11765. [PubMed: 16868085]
- Kitano H. Biological robustness. *Nat. Rev. Genet.* 2004; 5:826–837. [PubMed: 15520792]
- Lampe D, Churchill M, Robertson H. A purified Mariner transposase is sufficient to mediate transposition in vitro. *EMBO J.* 1996; 15:5470–5479. [PubMed: 8895590]
- Lampe DJ, Akerley BJ, Rubin EJ, Mekalanos JJ, Robertson HM. Hyperactive transposase mutants of the Himar1 Mariner transposon. *Proc. Natl. Acad. Sci. U.S.A.* 1999; 96:11428–11433. [PubMed: 10500193]
- Moore JP. A global view of epistasis. *Nat. Genet.* 2005; 37:13–14. [PubMed: 15624016]
- Moore D, Dowhan D. Purification and concentration of DNA from aqueous solutions. *Curr. Protoc. Mol. Biol.* 2002; 59:2.1.1–2.1.10.
- Ooi S, Shoemaker D, Boeke J. DNA helicase gene interaction network defined using synthetic lethality analyzed by microarray. *Nat. Genet.* 2003; 35:277–286. [PubMed: 14566339]
- Pan X, Yuan DS, Xiang D, Wang X, Sookhai-Mahadeo S, Bader JS, Hieter P, Spencer F, Boeke JD. A robust toolkit for functional profiling of the yeast genome. *Mol. Cell.* 2004; 16:487–496. [PubMed: 15525520]
- Pan X, Ye P, Yuan DS, Wang X, Bader JS, Boeke JD. A DNA integrity network in the yeast *Saccharomyces cerevisiae*. *Cell.* 2006; 124:1069–1081. [PubMed: 16487579]
- Parsons A, Lopez A, Givoni I, Williams D, Gray C, Porter J, Chua G, Sopko R, Brost R, Ho C. Exploring the mode-of-action of bioactive compounds by chemical-genetic profiling in yeast. *Cell.* 2006; 126:611–625. [PubMed: 16901791]
- Pozzi G, Masala L, Iannelli F, Manganelli R, Havarstein LS, Piccoli L, Simon D, Morrison DA. Competence for genetic transformation in encapsulated strains of *Streptococcus pneumoniae*: Two allelic variants of the peptide pheromone. *J. Bacteriol.* 1996; 178:6087–6090. [PubMed: 8830714]
- Roguev A, Bandyopadhyay S, Zofall M, Zhang K, Fischer T, Collins S, Qu H, Shales M, Park HO, Hayles J, Hoe KL, Kim DU, Ideker T, Grewal SI, Weissman J, Krogan N. Conservation and rewiring of functional modules revealed by an epistasis map in fission yeast. *Science.* 2008; 322:405–410. [PubMed: 18818364]
- Sassetti C, Boyd DH, Rubin EJ. Genes required for mycobacterial growth defined by high density mutagenesis. *Mol. Microbiol.* 2003; 48:77–84. [PubMed: 12657046]
- Schuldiner M, Collins S, Thompson N, Denic V, Bhamidpati A, Punna T, Ihmels J, Andrews B, Boone C, Greenblatt J. Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell.* 2005; 123:507–519. [PubMed: 16269340]

- St. Onge RP, Mani R, Oh J, Proctor M, Fung E, Davis R, Nislow C, Roth F, Giaever G. Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. *Nat. Genet.* 2007; 39:199–206. [PubMed: 17206143]
- Taubes G. The bacteria fight back. *Science.* 2008; 321:356–361. [PubMed: 18635788]
- Tong A. Global mapping of the yeast genetic interaction network. *Science.* 2004; 303:808–813. [PubMed: 14764870]
- Tong AH, Evangelista M, Parsons A, Xu H, Bader GD, Page N, Robinson M, Raghibizadeh S, Hogue CW, Bussey H, Andrews B, Tyers M, Boone C. Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science.* 2001; 294:2364–2368. [PubMed: 11743205]
- Tuomanen, L.; Mitchel, TJ.; Morrison, DA.; Spratt, BG. *The Pneumococcus.* ASM Press; Washington, D.C.: 2004.
- van Opijnen T, Bodi KL, Camilli A. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat. Methods.* 2009; 6:767–772. [PubMed: 19767758]
- Voytas D. Agarose gel electrophoresis. *Curr. Protoc. Mol. Biol.* 2000; 51:2.5A.1–2.5A.9.
- Wilmes GM, Bergkessel M, Bandyopadhyay S, Shales M, Braberg H, Cagney G, Collins S, Whitworth GB, Kress TL, Weissman J, Ideker T, Guthrie C, Krogan N. A genetic interaction map of RNA-processing factors reveals links between Sem1/DssI-containing complexes and mRNA export and splicing. *Mol. Cell.* 2008; 32:735–746. [PubMed: 19061648]
- Wilson K. Preparation of genomic DNA from bacteria. *Curr. Protoc. Mol. Biol.* 1997; 27:2.4.1–2.4.5.

Figure 1E.3.1.

A detailed schema of how the primers and adapter combine to result in a 120-bp DNA fragment that can be sequenced on an Illumina GAII. AP-B (AP-B_bc-ACAC) and AP-A (AP-A bc-ACAC) are two oligos that make up the adapter. AP-A has a random two base overhang (NN), which is employed to ligate the adapter onto the genomic DNA (*MmeI* leaves a random two-base overhang). Note that to prevent self-ligation, AP-B is slightly longer than AP-A. In addition, the adapter contains a 4-base barcode (bc) which enables mixing of different samples in a single flow cell lane. With primers P1-M6-MmeI and Gex PCR Primer 2 a 120-bp fragment is generated that contains Illumina specific sequences (the black tails with asterisk) that are necessary for sequencing. The Gex Sequencing primer is used for sequencing once samples are loaded on the flow cell (also see Tables 1E.3.1 and 1E.3.2 for oligo details).

Figure 1E.3.2.

Flowchart depicting Tn-seq, from library construction to massively parallel sequencing of transposon-chromosome junctions. **(A)** A gene-disruption library is constructed by transposing the mini-transposon *magellan6* into bacterial genomic DNA in vitro and subsequently transforming a bacterial population with the transposed DNA. The result is a pool of strains where each bacterium contains a single transposon randomly inserted in its genome. DNA is isolated from a portion of the bacterial pool (t_1); another portion is used to seed a culture on which selection is performed (in vitro or in vivo); after selection, bacteria are recovered and DNA is isolated again (t_2). **(B)** To accomplish sequencing of only the regions flanking *magellan6* insertions, DNA from both time points is digested with *MmeI*, which binds a sequence in the *magellan6* terminal inverted repeats but cuts 20 bp downstream, leaving a two-base overhang to which an adapter is ligated. A PCR is performed with one primer having complementarity to the adapter and the other complementary to the inverted repeat sequence. **(C)** The resulting PCR product is 120 bp long, with approximately 20 bp of bacterial specific DNA flanked by Illumina specific sequences needed for sequencing. After sequencing, different samples are split based on barcode sequence, and the bacteria-specific reads are mapped to the genome and counted for each insertion, thus allowing fitness to be calculated.

Figure 1E.3.3.

Tn-seq fitness details. **(A)** A gene's fitness is based on the average effect of multiple independent transposon insertions into a gene. On the x axis, the genomic region for *S. pneumoniae* TIGR4 is plotted covering genes SP0456 to SP0460. Black data points depict the fitness of a single transposon insertion within the coding region of a gene. A black line depicts the average fitness of a specific gene calculated from the individual insertions into that gene. For instance, the average fitness for genes SP0456, SP0458, and SP0460 is around 1 (same as wild type), while insertions into genes SP0457 and SP0459 are clearly disadvantageous and give average fitness values well below 1. The gray data points depict fitness from insertions into intergenic regions between genes. **(B)** Reproducibility between

independently generated libraries, consisting of 10,000 transposon insertion each is high ($R^2 = 0.92$). (C) Fitness for all genes across the *S. pneumoniae* genome.

Table 1E.3.1**Adapter Oligonucleotide Sequences**

Adapter	DNA sequence 5' to 3'	Purification
AP-A_bc-ACAC ^a	GTT CAG AGT TCT ACA GTC CGA CGA TCA <u>CAC</u> NN	PAGE
AP-B_bc-ACAC ^b	5Phos/ <u>GTG TGA</u> TCG TCG GAC TGT AGA ACT CTG AAC CTG TC/3Phos	PAGE

^aThe underlined sequence represents the barcode sequence (ACAC). The “NN” two-base random nucleotide overhang allows for ligation to the ends of *MmeI*-digested DNA fragments.

^bThe underlined sequence is the barcode sequence (GTGT; the complement of ACAC). The phosphorylated 5'-end allows for ligation to the dephosphorylated genomic DNA fragments. Note that AP-B_bc-ACAC is five nucleotides longer on the 3' end than AP-A_bc-ACAC, which prevents adapter-adapter ligation. In addition, the 3' phosphorylation of AP-B_bc-ACAC prevents the formation of adapter-adapter products.

Table 1E.3.2

PCR and Illumina Sequencing Primers

Primers	Primer sequence 5' to 3'	Purification
P1_M6_MmeI ^{ab}	CAA GCA GAA GAC GGC ATA CGA AGA CCG GGG ACT TAT CAT CCA ACC TGT	PAGE
Gex PCR Primer 2 ^b	AAT GAT ACG GCG ACC ACC GAC AGG TTC AGA GTT CTA CAG TCC GA	Standard desalting
Gex Sequencing Primer ^c	CGA CAG GTT CAG AGT TCT ACA GTC CGA CGA TC	Standard desalting

^aThe underlined sequence of primer P1-M6-MmeI is the complement of the Magellan inverted repeat.

^bThe italicized regions of primers P1-M6-MmeI and Gex PCR Primer 2 are Illumina specific sequences necessary for annealing to the oligos present in the flow cell.

^cThis primer is identical to the Illumina small RNA primer.