nature portfolio

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

Supplementary information

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

In the format provided by the authors and unedited

Supplementary Information

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

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

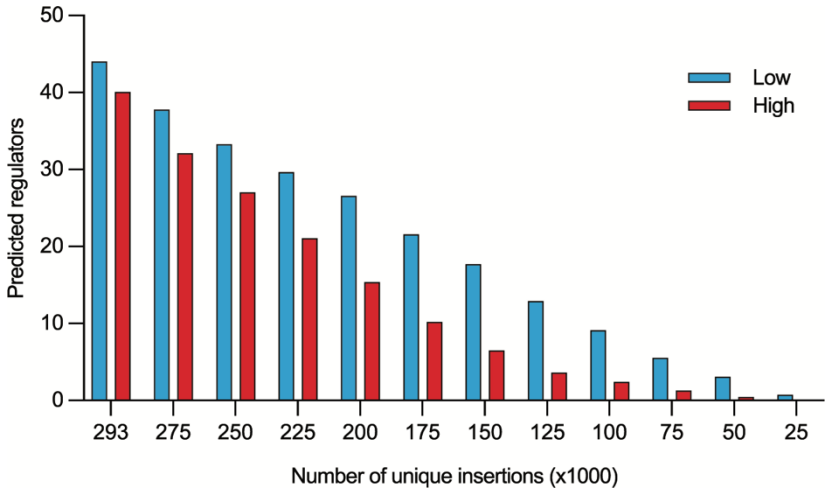
[1]Department of Microbiology and Immunology, University of Otago, PO Box 56, Dunedin 9054, New Zealand.

[2]Genetics Otago, University of Otago, Dunedin, New Zealand

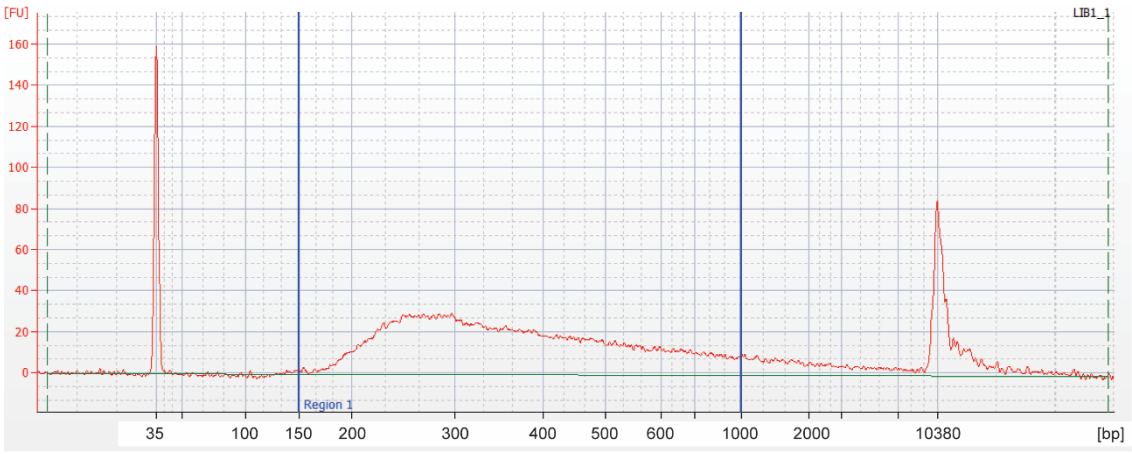
[3]Department of Biochemistry, University of Otago, PO Box 56, Dunedin 9054, New Zealand.

[4]Bio-Protection Research Centre, University of Otago, PO Box 56, Dunedin 9054, New Zealand.

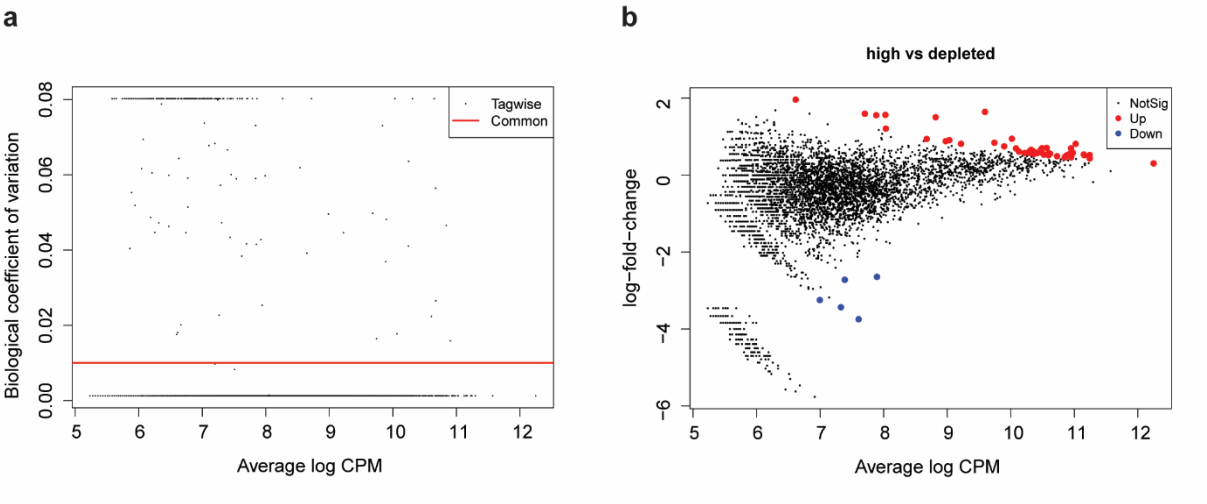
*For correspondence: peter.fineran@otago.ac.nz



Supplementary Figure 1. The number of regulators predicted from different transposon mutant insertion densities. Unique insertion counts from *csm* SorTn-seq were randomly subsampled to generate subsets representing different numbers of unique mutants (293,000 (total) to 25,000). The analysis was performed 50 times for each subsampled fraction, and data is presented as the mean number of significant predicted regulators (those having a $\log_2$ fold change > 0.5 and adjusted p. value < 0.05) in the low bin (blue) and high bin (red).



Supplementary Figure 2 Bioanalyzer trace of a sequencing library prepared during SorTn-seq. The large peaks at 35 and 10380 bp are the lower and upper markers, respectively. The blue lines at 150 and 1000 bp indicate the region designated (using the *region table* tool) to calculate average fragment size (395 bp).



Supplementary Figure 3. Dispersion and mean-difference plots generated by *edgeR*. a, Estimate of common and tag-wise dispersion for all libraries. **b,** Representative mean-difference (MD) plot, which shows $\log_2$ fold differences versus mean $\log_2$ counts per million (CPM) for each feature.

Supplementary Table 1. Mapping and insertion summary data for the *Serratia* chromosome. This table was modified from the .stats file output by the Bio-TraDIS[1] pipeline. Total reads are the number of reads following de-multiplexing. Reads with Tn*5* tag and % with Tn*5* tag are those containing the 12 nt transposon 'tag' generated during sequencing (2 mismatches allowed). Reads mapped and % mapped are those mapping to the *Serratia* LacA genome (NCBI reference sequence NZ_CP025085.1). Unique insertion sites (UIS) equals the total number of unique insertions. Length/UIS is the average number of nucleotides between unique insertions, as normalized by the size of the *Serratia* genome (4.97 MB). Table adapted from Smith et al., 2021[2].

Pool	Replicate	Total reads	Reads with Tn*5* tag	% with Tn*5* Tag	Reads mapped	% mapped	Unique insertion sites	Length/UIS
Low	1	305095	281902	92.40	228594	81.09	25113	197.98
High	1	272859	257771	94.47	212719	82.52	28828	172.46
Depleted	1	1119130	1047795	93.63	836287	79.81	113653	43.75
Low	2	496745	466994	94.01	389535	83.41	38411	129.44
High	2	384262	366482	95.37	288544	78.73	35826	138.78
Depleted	2	912755	849209	93.04	686601	80.85	135551	36.68
Low	3	392762	368103	93.72	322126	87.51	34831	142.74
High	3	378819	357311	94.32	315274	88.24	25148	197.70
Depleted	3	760684	722848	95.03	609174	84.27	115799	42.93
Input	1	1002283	953732	95.16	826990	86.71	129946	38.26
Input	2	1133121	1086507	95.89	892330	82.13	157246	31.62
Input	3	1062919	1009871	95.01	904636	89.58	124971	39.78

Supplementary Table 2. Mapping and insertion summary data for the pP1334 reporter plasmid. This table was modified from the .stats file output by the Bio-TraDIS[1] pipeline. Total reads are the number of reads following de-multiplexing. Reads with Tn*5* tag and % with Tn*5* tag are those containing the 12 nt transposon 'tag' generated during sequencing (2 mismatches allowed). Reads mapped and % mapped are those mapping to the pPF1334 reporter plasmid (Extended Data Figure 2). Unique insertion sites (UIS) equals the total number of unique insertions. Length/UIS is the average number of nucleotides between unique insertions, as normalized by the size of the pPF1334 reporter plasmid (6.25 KB).

Pool	Replicate	Total reads	Reads with Tn*5* tag	% with Tn*5* Tag	Reads mapped	% mapped	Unique insertion sites	Length/UIS
Low	1	305095	281902	92.40	1785	0.63	92	68.00
High	1	272859	257771	94.47	2001	0.78	118	53.02
Depleted	1	1119130	1047795	93.63	6824	0.65	406	15.41
Low	2	496745	466994	94.01	4735	1.01	168	37.24
High	2	384262	366482	95.37	6604	1.80	203	30.82
Depleted	2	912755	849209	93.04	5199	0.61	455	13.75
Low	3	392762	368103	93.72	737	0.20	53	118.04
High	3	378819	357311	94.32	2481	0.69	103	60.74
Depleted	3	760684	722848	95.03	2228	0.31	307	20.38
Input	1	1002283	953732	95.16	8687	0.91	638	9.81
Input	2	1133121	1086507	95.89	7658	0.70	696	8.99
Input	3	1062919	1009871	95.01	4030	0.40	420	14.90

Supplementary Table 3. Total colony forming units for *E. coli* donor and *Serratia* recipients. Average (avg) counts shown are from triplicate dilution/plating. CFU; colony forming units, SD; standard deviation, ALA; 5-aminolevulinic acid hydrochloride, Cm; chloramphenicol.

Pre-conjugation (mating mixture)					
Selection used (strain)	**Replicate 1**	**Replicate 2**	**Replicate 3**	**Total CFU for mating mixture (avg absolute CFU)**	**SD**
ALA (*E. coli*)	2.50×10^8	2.30×10^8	2.80×10^8	2.53×10^8	2.05×10^7
Cm (*Serratia* w/ reporter)	4.20×10^8	3.10×10^8	3.90×10^8	3.73×10^8	4.64×10^7

Supplementary Table 4. *E. coli* and *Serratia* counts following transposon mutagenesis. Average (avg) counts shown are from triplicate dilution/plating. CFU; colony forming units, SD; standard deviation, tn; transposon, ALA; 5-aminolevulinic acid hydrochloride, Cm; chloramphenicol, Km; kanamycin.

Pre-selection library (6h)							
Selection used (strain)	**Replicate 1**	**Replicate 2**	**Replicate 3**	**CFU/ml of resuspended mating spots (avg)**	**SD**	**Total CFU recovered (11ml)**	**Estimated unique tn insertions in library***
ALA (*E. coli*)	2.50×10^8	2.20×10^8	2.80×10^8	2.50×10^8	2.45×10^7	2.75×10^9	
Cm (*Serratia* w/ reporter)	1.80×10^8	1.60×10^8	1.40×10^8	1.60×10^8	1.63×10^7	1.76×10^9	
Km (*Serratia* w/ tn)	1.50×10^5	1.60×10^5	2.10×10^5	1.73×10^5	2.62×10^4	1.91×10^6	
Cm + Km (*Serratia* w/ reporter and tn)	2.30×10^5	2.40×10^5	2.80×10^5	2.50×10^5	2.16×10^4	2.75×10^6	**5.83×10^5**

*Estimated based on the total number of recovered transposon mutants normalized by the fold change in total *Serratia* counts pre-conjugation to pre-selection (6h).

Supplementary Table 5. *E. coli* and *Serratia* counts following library outgrowth. Average (avg) counts shown are from triplicate dilution/plating. CFU; colony forming units, SD; standard deviation, tn; transposon, ALA; 5-aminolevulinic acid hydrochloride, Cm; chloramphenicol, Km; kanamycin.

Post-Outgrowth (24h)					
Selection used (strain)	**Replicate 1**	**Replicate 2**	**Replicate 3**	**CFU/ml of outgrowth (avg)**	**SD**
ALA (*E. coli*)	4.20×10^5	3.50×10^5	4.00×10^5	3.90×10^5	2.94×10^4
Cm (*Serratia* w/ reporter)	2.00×10^8	2.10×10^8	2.30×10^8	2.13×10^8	1.25×10^7
Km (*Serratia* w/ tn)	1.80×10^8	2.90×10^8	2.30×10^8	2.33×10^8	4.50×10^7
Cm + Km (*Serratia* w/ reporter and tn)	2.90×10^8	1.30×10^8	2.80×10^8	2.33×10^8	7.32×10^7

Supplementary Table 6. Strains used in this study.

Strain Name	Description	Notes	Reference
Escherichia coli			
ST18	Auxotrophic donor for biparental conjugation	Sm^R, requires ALA	(Thoma and Schobert, 2009)[3]
***Serratia* sp. ATCC 39006**			
PCF396	Δ*pigA-O* (RS20700-RS20630)	Pigmentless wild type	(Smith et al., 2021)[2]

Supplementary Table 7. Plasmids used in this study. Multiple cloning site (MCS); origin of replication (Ori); origin of transfer (OriT).

Name	Description	Features	Reference
Transposon mutagenesis			
pKRCPN2	Delivery vector containing transposase and transposon carrying a *uidA* (β-glucuronidase) reporter gene and kanamycin resistance cassette	R6K/Ori, RP4/OriT, Tn-DS1028*uidA* Km, Transposase, Tc^R	(Monson *et al*., 2015; Patterson *et al*., 2015)[4,5]
Fluorescent reporters			
pPF1307	*Pcsm-eYFP* (type III-A system reporter)	pBR322/Ori, RP4/OriT, Cm^R, MCS, eYFP	(Smith et al., 2021)[2]
pPF1334	*Pcsm-eYFP* (type III-A system reporter) + IPTG-inducible *mCherry*	pBR322/Ori, RP4/OriT, Cm^R, lacI/T5, MCS, mCherry, eYFP	(Smith et al., 2021)[2]
pPF1438	IPTG-inducible *mCherry* only vector	pBR322/Ori, RP4/OriT, Cm^R, lacI/T5, mCherry	(Smith et al., 2021)[2]

Supplementary Table 8. Configuration of the FACSAria Fusion cell sorter using during SorTn-seq^2. The cell sorter is equipped with four lasers that excite at different wavelengths. Mirrors and bandpass filters direct optical signals towards the appropriate electronic detectors.

Laser (excitation wavelength)	Mirror	Bandpass filter
Blue (488 nm)	502	530/30
	655	695/40
Green (561 nm)	-	582/15
	600	610/20
	630	670/14
	685	710/50
	735	780/60
Red (640 nm)	-	670/30
	690	730/45
	755	780/60
Violet (405 nm)	-	450/50
	505	525/50
	595	610/20
	630	660/20
	690	710/50
	750	780/60

Read trimming using PRINSEQ

Timing: 20 min, 10 min hands-on

NOTE The following protocol is used to trim reads to a specified length and is performed in a terminal window. FASTQ files used during SorTn-Seq analysis[2] were trimmed from the 3' end to a final length of 50 nt using PRINSEQ. **CRITICAL** Do not trim reads from the 5' end as this will remove the transposon 'tag' sequence.

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

```
perl /APPS/prinseq-lite-0.20.4/prinseq-lite.pl -trim_to_len 50 -fastq filename.fastq
```

Parameter*	Action
`-trim_to_len`	Specifies the final read size generated through 3' trimming
`-fastq`	Specifies the FASTQ file to be trimmed

*For a full description of PRINSEQ parameters execute the command `perl prinseq-lite.pl -h` on the command line.

References

1 Barquist, L. *et al.* The TraDIS toolkit: sequencing and analysis for dense transposon mutant libraries. *Bioinformatics* **32**, 1109-1111, doi:10.1093/bioinformatics/btw022 (2016).

2 Smith, L. M. *et al.* The Rcs stress response inversely controls surface and CRISPR-Cas adaptive immunity to discriminate plasmids and phages. *Nat Microbiol* **6**, 162-172, doi:10.1038/s41564-020-00822-7 (2021).

3 Thoma, S. & Schobert, M. An improved *Escherichia coli* donor strain for diparental mating. *FEMS Microbiol. Lett.* **294**, 127-132 (2009).

4 Monson, R. *et al.* A Plasmid-Transposon Hybrid Mutagenesis System Effective in a Broad Range of Enterobacteria. *Front Microbiol* **6**, 1442, doi:10.3389/fmicb.2015.01442 (2015).

5 Patterson, A. G., Chang, J. T., Taylor, C. & Fineran, P. C. Regulation of the Type I-F CRISPR-Cas system by CRP-cAMP and GalM controls spacer acquisition and interference. *Nucleic Acids Res.* **43**, 6038-6048, doi:10.1093/nar/gkv517 (2015).