

Pooled Assembly, Genotyping and Scoring of Complex Synthetic Genomic Libraries

*Robert Egbert^{1,2} (rgegbert@lbl.gov), Eric Yu², Dylan McCormick³, Ben Adler², **Adam Arkin**^{1,2}

¹Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA; ²Department of Bioengineering, University of California Berkeley; ³Department of Molecular and Cell Biology, University of California Berkeley.

Precision engineering of complex behaviors in living systems is complicated by our limited understanding of the contextual determinants of gene expression and our limited ability to manipulate the genetic code in individual cells at scale. Iterative design, build and test cycles thus often sample only a small fraction of the functional parameter space. Further, the relative ease of optimizing gene networks on plasmids for bacterial expression often is not predictive of system performance when chromosomally expressed for deployment in complex environments from bioreactor to soil or gut microbial communities. We have developed a comprehensive engineering platform for pooled assembly, genotyping and single-variant fitness scoring of barcoded genomic libraries coupled with retrieval of individual genotypes from the library via CRISPRi. Mapping individual genotypes to fitness drives predictions of function across the sequence landscape, balancing the costs and benefits of expression.

To validate the genome engineering platform we generated a genomic library of over two million barcoded variants of the violacein biosynthetic pathway *vioABCDE* from *Chromobacterium violaceum* in *E. coli*, sampling from a genotype space of over 260,000 combinations. We employed λ -Red recombination to serially integrate five DNA fragment libraries - here, each a gene with a degenerate ribosome binding site - fused to a selection-enrichment cassette (SEC) with homology for the target genomic locus. Successive “inchworm” integration stages replaced the SEC from the previous stage, cycling among SECs, each consisting of one of two fluorescence markers and one of three antibiotic resistance markers. We pooled transformants from each integration stage using fluorescence activated cell sorting to enrich for the expected fluorescence phenotype and screen out spontaneous resistance mutants and off-target integration events manifest as cells with errant fluorescent phenotypes.

We genotyped the pooled library by associating an individual barcodes with the ribosome binding site sequence of each *vio* gene by deep sequencing amplicon fusion libraries generated by emulsion PCR¹. Time-series measurements of barcode

abundance allow us to generate fitness scores for each barcoded variant under pathway induction conditions and in resource competition with violacein-sensitive *Bacillus subtilis* to identify variants that optimally balance the benefit of violacein production as an antimicrobial against the cost of gene expression. We are analyzing the genotypes and fitness scores to comprehensively map the expression space and predict sequence to function relationships.

To isolate individual genotypes from the pooled library we developed a barcode-specific CRISPR interference (CRISPRi) technique to use with fluorescence-activated cell sorting. We used the random barcode embedded in the SEC of the final integration stage to repress GFP expression with the RNA-encoded guide sequence specific repressor dCas9². We validated the isolation of individual genotypes from a library of eight clones with a wide range of GFP expression levels by sorting the library transformed with CRISPRi plasmids that encode dCas9 with a guide RNA that target individual barcodes. We have also demonstrated the retrieval of low-abundance cell types from the pooled genomic library by transforming the library with CRISPRi plasmids encoding guides that target rare barcodes and sorting for CRISPRi-responsive cells.

We believe our approach to high-throughput construction, analysis and manipulation of pooled genomic libraries through comprehensively sampling of the parameter space of engineered gene networks will aid the prediction of sequence to function relationships and enable complex engineered phenotypes that are unreachable with existing methods. We anticipate this genomic assembly, screening and isolation platform will advance synthetic biology efforts to optimize large biosynthetic gene clusters for expression of natural products and engineering of other complex cellular behaviors encoded by multiple genes.

This work is supported by the Genome Science program within the Office of Biological and Environmental Research (Project grant number DE-SC008812, Funding Opportunity Announcement DE-FOA-0000640)

References

1. Zeitoun, R.I. *et al.* 2015. Multiplexed tracking of combinatorial genomic mutations in engineered cell populations. *Nature Biotechnol.* **33**, 6, 631–637.
2. Qi, L. S. *et al.* 2013. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183.