109. Tracking Combinatorial Engineered (TRACE) libraries at the genome-scale

Ramsey I. Zeitoun[1*] (ramsey.zeitoun@colorado.edu), Andrew D. Garst[1], George D. Degen[2], **Ryan T. Gill[1]**

[1]University of Colorado Boulder; [2]Carnegie Mellon University

Project Goals

The ability to perform genome engineering and editing on practical timescales requires DNA target enrichment and sequencing methods that can be used to quantitatively interrogate libraries containing billions of combinatorially engineered mutants.[1-3] Current library construction methods for genome-engineering operate at throughputs of billions per week, while tracking methods are limited to scales of 100---1000s per week or to operating in a non-combinatorial manner. The result is an inability to complete design-build-test-learn cycles at the overall throughputs required for combinatorial genome engineering (~10^8/week).

The major challenge to the development of combinatorial tracking methods is the necessity to identify mutations located at distant locations throughout a microbial genome, and to do so at high throughput and on a cell specific basis. Current methods for genotyping distant genomic sites are relatively low throughput and can be low resolution. For example, multiplexed allele specific PCR (MASC PCR) reports on the presence or absence of the wild--- type allele in single clones rather than the actual sequence of interest, and is typically limited to throughputs of 100-1000s of clones because of the need to perform individual PCR reactions on each clone.[4-6] Next-gen sequencing approaches are inexpensive, high-throughput, and high-resolution, but suffer from read lengths (several 100 nt) that are too short to allow identification of distantly located mutations, and thus also require separation and tagging strategies that reduce throughput in a manner similar to MASC PCR. Single cell sequencing approaches can be used to solve this problem, but these approaches currently do not operate at the throughputs and depth required.[7] Ideally, combinatorial tracking technology will allow one to interrogate genomes at dozens of different locations and provide quantitative data on individual combinations of mutants across combinatorial libraries containing billions of clones.

Linking PCR offers a possible genotyping solution as this method can be applied to stitch together PCR products from distal chromosomal origins in a single pot reaction. Not only does this significantly reduce the costs associated with colony-based Sanger sequencing by reducing the number of sequencing reactions per clone, but combining this method with emulsification technologies (Linking-emulsification PCR or LE-PCR) offers a potentially useful route to multiplexed genotyping of large populations (Figure 1a). The inefficiency of assembly and amplification in LE-PCR however has thus far limited its application to the study of only two distinct sites.8 To be widely applicable as a technology for tracking targeted genomic libraries, such an approach would ideally enable sampling of many genomic sites. To expand the number of gene targets that can be tracked and improve the overall efficiency of this approach requires an automated de novo primer design to optimize the many criteria necessary for efficient PCR based assembly and subsequent analysis. We thus sought to automated optimization of both the primer design and reaction conditions. This method allows us to achieve assembly of a larger number of genomic sites as proof of concept for multiplexed genotyping of distal chromosomal sites of interest. We then applied these conditions towards Tracking Combinatorial Engineered libraries by coupling the linking chemistry to emulsion PCR and resolving the products using next-generation sequencing technologies.

References

1. Wang, Harris H., et al. "Programming cells by multiplex genome engineering and accelerated evolution." *Nature* 460.7257 (2009): 894-898.
2. Sandoval, Nicholas R., et al. "Strategy for directing combinatorial genome engineering in Escherichia coli." *Proceedings of the National Academy of Sciences* 109.26 (2012): 10540-10545.
3. Warner, Joseph R., et al. "Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides." *Nature biotechnology* 28.8 (2010): 856-862.
4. Cha, Rita S., et al. "Mismatch amplification mutation assay (MAMA): application to the cH-ras gene." *Genome Research* 2.1 (1992): 14-20.
5. Wang, Harris H., and George M. Church. "Multiplexed genome engineering and genotyping methods applications for synthetic biology and metabolic engineering." *Methods Enzymol* 498 (2011): 409-426.
6. Carr, Peter A., et al. "Enhanced multiplex genome engineering through co-operative oligonucleotide co-selection." *Nucleic acids research* 40.17 (2012): e132---e132.
7. Wang, Daojing, and Steven Bodovitz. "Single cell analysis: the new frontier in 'omics'." *Trends in biotechnology* 28.6 (2010): 281---290.
8. Wetmur, James G., et al. "Molecular haplotyping by linking emulsion PCR: analysis of paraoxonase 1 haplotypes and phenotypes." *Nucleic acids research* 33.8 (2005): 2615-2619.