High-throughput pooled functional screens *via* in-vivo production of Single-stranded DNA, toward applications in photosynthetic microbial hosts

Max G. Schubert[1]* (max_schubert@hms.harvard.edu), Daniel B. Goodman[2], Fahim Farzadfard,[3] Timothy K. Lu,[3] Seth L. Shipman[2], Himadri Pakrasi[4],Pramod Wangikar[5], **George M. Church**[1,3]

[1]Harvard University, Boston, MA; [2]University of California at San Francisco, CA; [3]Massachussetts Institute of Technology, Cambridge, MA; [4]Washington University, St. Louis, MO; [5]IIT-Bombay, Powai, Mumbai, India

http://arep.med.harvard.edu

Project Goals: Enable and explore recombineering using substrates produced *in-vivo*, via specialized bacterial retro-elements, creating a pooled Functional Genomics system, combining efficient editing and NGS-based tracking of mutants. Deploy this tool and other Next-generation genome editing technology toward Energy-related goals, including the study of photosynthesis.

Tremendous genetic variation exists in nature, but our ability to create and characterize individual genetic variants remains far more limited in scale. Likewise, synthetic variants aid our understanding of gene and genome function, but computational design of variants outpaces experimental measurement of their effect. Here, we show *in-vivo* production of single-stranded DNA via the targeted reverse-transcription of Retrons enables efficient and continuous generation of precise genomic edits in *Escherichia coli* at greater than 90% efficiency[1]. Curiously, the newly-identified single-stranded annealing protein (SSAP) CspRecT improves efficiency of this process by as much as 11-fold, a larger effect than observed when CspRecT is used with electroporated DNA. Because barcoded mutations are created, this tool also effectively couples phenotypes to a targeted sequencing output, enabling pooled high-throughput screens of genetic variants, a process we call Retron Library Recombineering (RLR)[1]. We measure antibiotic resistance resulting from synthetic variants using both qualitative and quantitative RLR protocols for pooled phenotypic measurement. RLR can also be performed using natural variants as input, and we demonstrate this by using sheared genomic DNA of an evolved bacterium as an input substrate for RLR. In this way, we identify causal variants leading to antibiotic resistance, and demonstrate a saturating genomic RLR library, in which tens of millions of barcoded experiments are performed within each single pool, and all genetic variants in a strain are exhaustively tested. The capacity to use non-designed DNA for such a screen stands in contrast to CRISPR-based methods, and opens the door to many applications using undefined and/or degenerate variation. Pooled genomic editing using ssDNA produced *in vivo* thus presents new avenues for creating and exploring variation at the whole genome scale.

The future of Bioenergy depends on applying these rapid genetic engineering techniques to organisms using atmospheric CO_2 for carbon, and sunlight or renewables for energy. To this end, we have begin to apply our approaches to fast-growing cyanobacteria, an exciting new area of bioenergy research[2,3]. Growing at speeds previously thought impossible, these photosynthetic bacteria are a possible next-generation renewable synthetic biology host. Because functional genomics approaches depend on complete, accurate genomes, we have begun by completing the genome of PCC11801, the cyanobacterial strain with the fastest growth rates in ambient air (~2.3hr doubling time)[3]. Our draft genome resolves existing assembly gaps, establishes that PCC11801 is a more distant relative to the model strain PCC7942 than was previously

appreciated, and confidently detects 6 novel plasmids stably maintained in this strain. These previously unreported plasmids represent ~12% of the genetic material in the strain, possibly important for the strain's intriguing phenotypes. Each plasmid shares locally collinear blocks with other plasmids reported in this *Synechococcus* clade, suggesting an episomal component to the core genome of *Synechococcus*. Directed evolution in this strain background enabled us to isolate a strain growing to higher density in batch growth, and having less adherence to the growth vessel. We are excited to develop pooled functional genomics approaches for fast-growing cyanobacteria, and our highest priority is to use our existing approaches[4-6] to develop recombineering technologies in this organism. These technologies could aid our investigate these new phenomena, and further enable this new renewable synthetic biology chassis as they have for past hosts[6].

References

1. Schubert, M. G. et al. High throughput functional variant screens via in-vivo production of single-stranded DNA. Cold Spring Harbor Laboratory 2020.03.05.975441 (2020) doi:10.1101/2020.03.05.975441.

2. Yu, J. et al. Synechococcus elongatus UTEX 2973, a fast growing cyanobacterial chassis for biosynthesis using light and CO_2. Sci. Rep. 5, 8132 (2015).

3. Jaiswal, D. et al. Genome Features and Biochemical Characteristics of a Robust, Fast Growing and Naturally Transformable Cyanobacterium Synechococcus elongatus PCC 11801 Isolated from India. Sci. Rep. 8, 16632 (2018).

4. Filsinger, G. T. et al. Characterizing the portability of phage-encoded homologous recombination proteins. Nat. Chem. Biol. (2021) doi:10.1038/s41589-020-00710-5.

5. Wannier, T. M. et al. Improved bacterial recombineering by parallelized protein discovery. Proc. Natl. Acad. Sci. U. S. A. 117, 13689–13698 (2020).

6. Wannier, T. M. et al. Recombineering and MAGE. Nature Reviews Methods Primers 1, 1–24 (2021).

This project has been funded by DOE grant DE-FG02-02ER63445. Dr. Church is a founder of companies in which he has related financial interests: ReadCoor; EnEvolv; and 64-x. For a complete list of Dr. Church's financial interests, see also arep.med.harvard.edu/gmc/tech.html.