Experimental and Computational Tools for Sequence to Activity Mapping

Gur Pines* (gur.pines@colorado.edu), Andrew Garst, Marcelo Bassalo, Andrea L. Halweg-Edwards and **Ryan T Gill**

University of Colorado, Boulder

http://www.dynamcc.com

Project Goals: A comprehensive sequence to activity mapping is central to a successful protein, pathway and genome level engineering. Even at the single protein level, the mutational space is too vast for complete exploration. We develop a codon compression approach that helps in reducing library size when performing saturation mutagenesis using degenerate oligonucleotides. We also develop a method to systematically edit genomic loci in the single codon resolution. This method, termed CRISPR EnAbled Trackable genome Engineering (CREATE) combines traditional lambda red recombineering with the high efficiency editing of the CRISPR-CAS9 system. Using CREATE, we successfully saturated every site in the essential E. coli *folA* gene and mapped both the essential sites for activity and specific mutations conferring resistance to its inhibitor, trimethoprim.

The common approach for saturation mutagenesis is using NNK (N = A/C/G/T, K = G/T) which covers all amino acids in a single codon. However, since NNK codes for 32 codons, some of the 20 amino acids are represented more than once, leading to amino acid bias. Moreover, this approach will always include the wild type (wt) amino acid and a stop codon. These factors needlessly increase library size and downstream screening efforts. Also, depending on the target organism, some of the NNK codons might suffer from extremely low usage hindering protein expression. Our computational approach allows to define the exact amino acid collection that is desired, along with the target organism. The output is a pool of degenerate codons that covers the exact query without any off target amino acids, and with high usage codons, as defined by the user[1]. For example, when completely saturating a site, the compressed codons will cover 19 amino acids, excluding the wt and stop codons. Another option is to define an exact set of amino acids to replace the wt. This can be knowledge-based or in order to keep the wt residue properties such as hydrophobicity or size. Another feature allows to define the level of redundancy within the codon pool. We added this function since it is known that in some cases synonymous mutations may have a significant effect on mRNA stability, protein folding rates and more[2,3]. We now have the codon compression tools in a dedicated website to allow easy access. The website also allows to upload custom genetic and usage tables enabling users that are interested in non standard or synthetic genetic codes to use these tools. We hope that this

improved access will increase the audience to our approach and will help in speeding up protein engineering.

CREATE is a method for rapid and efficient genome editing in the single codon resolution over tens to hundreds of thousands of loci in parallel. This method combines phage-based recombineering[4] with CRISPR genome editing[5]: site-specific gRNAs and editing templates are synthesized in parallel on a microarray and later pool-cloned into a plasmid. The gRNA, together with an inducible Cas9 digest the genome in a sequence-specific manner, while the editing template is integrated genomically by the recombineering machinery. The editing template harbors two mutations: the first is the desired modification and the second silently mutates a PAM sequence, which is necessary for CAS9 cleavage. This approach allows to build genome-wide libraries and complete saturation of whole genes. Since the plasmid editing templates highly correlates with the genomic edits, simple plasmid sequencing replaces whole genome interrogation, making this system easily trackable. We demonstrate CREATE by completely saturating the essential E. coli *folA* gene, generating its full sequence to activity map. Moreover, we also challenged these library cells with trimethoprim, a FolA specific inhibitor, and found two mutations within the same site that confer resistance to this drug.

References

1. Pines, G. *et al.* Codon compression algorithms for saturation mutagenesis. *ACS Synth. Biol.* **4,** 604–614 (2015).
2. Plotkin, J. B. & Kudla, G. Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.* **12,** 32–42 (2011).
3. Firnberg, E., Labonte, J. W., Gray, J. J. & Ostermeier, M. A Comprehensive, High-Resolution Map of a Gene's Fitness Landscape. *Mol. Biol. Evol.* **31,** 1581–1592 (2014).
4. Pines, G., Freed, E. F., Winkler, J. D. & Gill, R. T. Bacterial Recombineering: Genome Engineering via Phage-Based Homologous Recombination. *ACS Synth. Biol.* **4,** 1176–1185 (2015).
5. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31,** 233–239 (2013).

This work is supported by the Office of Biological and Environmental Research in the DOE Office of Science.