A workflow for the systems-level analysis, design, and engineering of genomically recoded organisms

Anush Chiappino-Pepe[1,2]* (anush_chiappinopepe@hms.harvard.edu), Lisa Dratva[1,3], Alexandra Rudolph[1], and **George M. Church**[1,2]

[1]Department of Genetics, Harvard Medical School, Boston, MA, USA; [2]Wyss Institute for Biologically Inspired Engineering, Boston, MA, USA; [3]School of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

http://arep.med.harvard.edu

We are establishing a computational workflow for the phenotypic analysis, design, and engineering of genomically recoded organisms. This workflow connects genome sequences to cellular phenotypes with improved nucleotide resolution by relying on a set of identified genome design rules. For this purpose, we have developed deep learning models connecting sequence-to-function of genetic features; extended genome-scale models of metabolism, expression, and regulation that predict cellular fitness from genome sequences; and methods for the analysis of recoding mutations.

Genomically recoded organisms are tightly biocontained and biosolated (*e.g.*, virus resistant) and allow the efficient incorporation of multiple non-standard amino acids, making them attractive platform technologies for biotechnological, industrial, and biomedical applications[1–4]. In the process of recoding, we substitute a set of selected codons by synonymous ones throughout the entire genome[2,5]. Although protein sequences are maintained, we and others have observed considerable fitness reductions upon recoding[2,3,6,7]. To design viable genomically recoded organisms, we need computational methods that can connect genome sequences to cellular phenotypes with enhanced resolution than available methods.

Here, we present a workflow for the rational analysis, design, and engineering of genomically recoded organisms. First, we have developed deep learning models connecting sequence-to-function of genetic features that are mechanistically intractable. Second, we have constructed extended genome-scale models for the integrative analysis of metabolism, expression, and regulation. These models can predict cellular fitness from recoded genome sequences and can integrate large and disparate omics data sets. Finally, we have defined optimization-based methods for the computational analysis of recoding mutations. With this workflow, we can predict at a systems-level the combinatorial effect of recoding mutations on fitness, and we identify sets of mutational bottlenecks constraining growth.

We are applying this workflow to identify fitness-decreasing mutations in a 57-codon *Escherichia coli* genome[6]. In parallel, the models developed together with newly obtained RNA-seq data for strains containing segments of a 57-codon *Escherichia coli* genome are helping characterize the metabolic function arising from synthetic genome segments.

In the future, this workflow will serve to design genomically recoded organisms with minimal growth reduction and will provide a better understanding of genotype-to-phenotype relationships at a cellular level.

References/Publications

1. Ostrov, N. *et al.* Synthetic genomes with altered genetic codes. *Curr Opin Sys Biol* **24**, 32–40 (2020).
2. Lajoie, M. J. *et al.* Genomically Recoded Organisms Expand Biological Functions. *Science* **342**, 357–360 (2013).
3. Fredens, J. *et al.* Total synthesis of Escherichia coli with a recoded genome. *Nature* **569**, 514–518 (2019).
4. Robertson Wesley E. *et al.* Sense codon reassignment enables viral resistance and encoded polymer synthesis. *Science* **372**, 1057–1062 (2021).
5. Isaacs, F. J. *et al.* Precise Manipulation of Chromosomes in Vivo Enables Genome-Wide Codon Replacement. *Science* **333**, 348 (2011).
6. Ostrov, N. *et al.* Design, synthesis, and testing toward a 57-codon genome. *Science* **353**, 819–822 (2016).
7. Lajoie, M. J. *et al.* Probing the Limits of Genetic Recoding in Essential Genes. *Science* **342**, 361 (2013).

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-FG02-02ER63445. Dr. Church is a founder of companies in which he has related financial interests: GRO Biosciences, ReadCoor; EnEvolv; and 64-x Bio. For a complete list of Dr. Church's financial interests, see also arep.med.harvard.edu/gmc/tech.html.