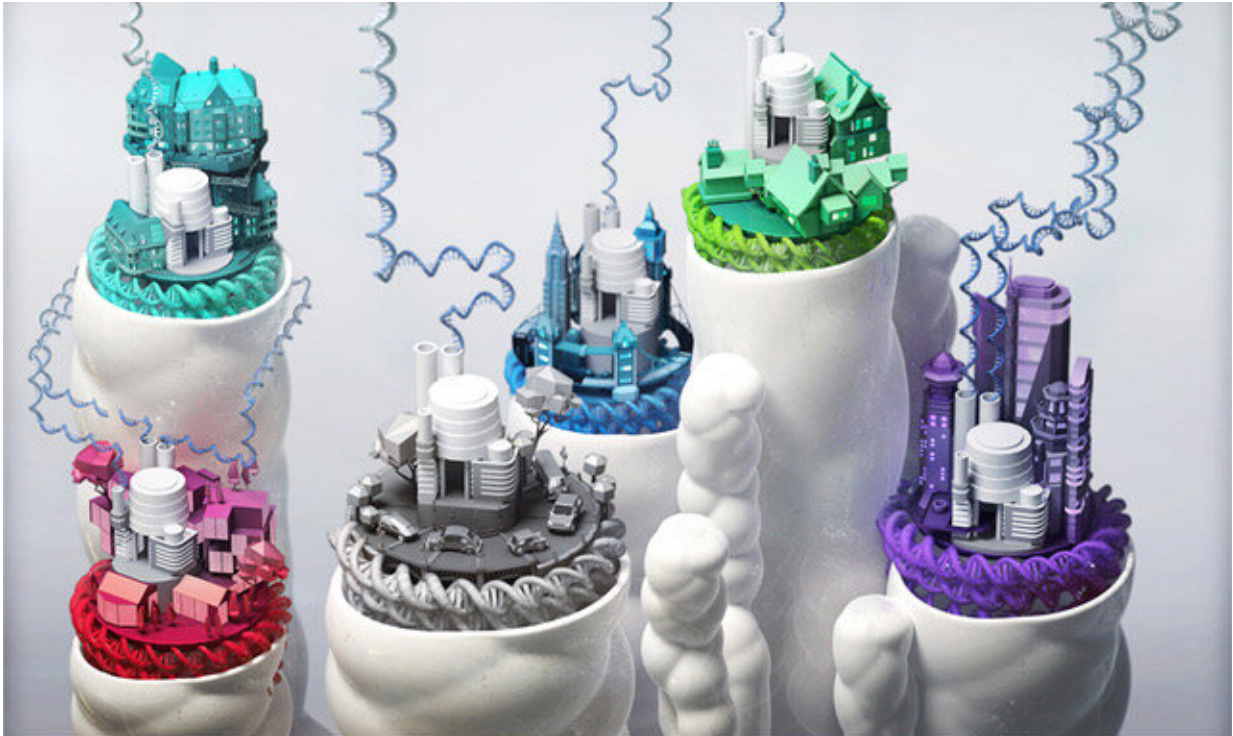


# Understanding genomes, piece by piece

March 3 2022

---



Gene expression is influenced by neighboring transcriptional patterns. Genetic rearrangements in a synthetic yeast chromosome revealed that characteristics of a gene's new transcriptional neighborhood predicted its transcript isoform boundaries and abundance. Credit: Tobias Wüstefeld

The celebrated physicist Richard Feynman is credited with the quote, "What I cannot create, I do not understand." As well as informing Feynman's approach to theoretical physics, it's a good way of describing the motivations of synthetic biologists, with their interest in building

genomes from scratch. By designing and building synthetic genomes, they hope to better understand the code of life. Synthetic biology has been organized around the concept of using DNA sequences as 'parts' with reproducible functions. Now, through successful collaborations and the use of cutting-edge tools, EMBL's Steinmetz Group has gained an important insight into the variation of gene expression that results from the position or context of these DNA parts within the genome.

Explaining the underlying question motivating the work, Amanda Hughes—co-lead author and postdoc in the Steinmetz Group—said, "In [synthetic biology](#), you tend to break things down into modular, 'plug-and-play' parts. These are promotor parts, coding regions, and terminator parts. We wanted to test whether these pieces really are 'plug-and-play,' functioning the same way in any context, or whether their position affects their function. We wanted to better understand how the linear organization of genes affects their functions and identify general design principles that could be applied to building genomes."

## **A synthetic biology toolbox delivers contextual insights**

This work, funded by BMBF and Volkswagen Foundation's "Life?" initiative, was possible because of two key technologies: synthetic yeast strains from the Sc2.0 consortium and long-read direct RNA sequencing. The strains obtained from the Sc2.0 consortium included a design feature called 'SCRaMbLE' that provides the ability to rearrange genes into different locations at a previously unachievable scale. The expertise and tools available in the Genomics Core facility at EMBL, including Oxford Nanopore's GridION, allowed the team to perform long-read direct RNA sequencing, permitting identification of both the start and end of RNA molecules and their assignment to particular rearrangements. The combination of these cutting-edge technologies was

critical to measure full-length RNA molecules from genes across many contexts.

The paper, published in *Science* showed that context—and in particular transcriptional context—alters the RNA output of a gene. Using long-read direct RNA sequencing, they were able to observe changes in the start, end, and amount of full-length RNA molecules expressed from DNA sequences that had been randomly rearranged in synthetic yeast genomes. Relocating a gene affected the length and abundance of its RNA output; however, these changes were not always explained by the new adjacent DNA sequence. It appeared to be transcription occurring around it, rather than the sequence itself, that altered a gene's RNA output.

Gleaning general principles from such a large, stochastic dataset was not a trivial task, as the lead author Aaron Brooks explained that "to reach our conclusions, we had to observe genes in many alternative genetic contexts, which were present in the SCRaMbLE strains. Putting the pieces back together, however, was a big effort. We had to generate a massive sequencing dataset, which, in turn, required us to develop new software tools. We had to rely on sophisticated machine learning algorithms to help us understand the complex patterns we were observing." Modeling a gene's RNA output based on its new upstream and downstream contexts revealed that features related to surrounding transcriptional patterns predicted RNA boundaries and abundance. For example, if a gene was relocated next to a highly expressed neighbor, its expression also tended to increase.

## **Defining design principles for building genomes**

In addition to illuminating the relationship between RNA abundance and neighboring gene expression, the researchers also noted a compelling relationship between the end positions of RNAs of convergent genes

(genes oriented with ends towards one another). Specifically, they found that the length of an RNA was affected by the proximity and abundance of neighboring transcripts. Jef Boeke, co-author and Director of the Sc2.0 consortium remarked on these insights that "deep transcriptional profiling combined with the genomic variations produced using [the] SCRaMbLE system have given us new insights into the flexibility of the yeast [genome](#) and pointed out that the rules of where transcripts end can be surprisingly context dependent."

Ultimately, applying these findings, the researchers were able to tune the length of RNA molecules by controlling transcription of a neighboring gene. The team demonstrated that the lessons learned from studying the transcriptomes of SCRaMbLEd genomes can be applied to engineer genomes with desired functions. The study also proposes a new synthetic biology design concept that the researchers term 'transcriptional embedding' that could be used to reversibly tag an RNA, altering its stability, translation into protein, or even localisation. All of this could be accomplished, they believe, by controlling the expression of a convergent, neighboring gene rather than the gene itself.

"The unbiased and high-throughput nature of the gene reshuffling approach used here leads us to discover functions of genomic sequences in different genomic contexts, something that previously was not possible at scale," said Lars Steinmetz, group leader at EMBL. "This approach emphasizes that context matters in regulating transcript ends—surprisingly, even permitting context-dependent predictions of transcript ends when genes are reshuffled to new locations. Ultimately, the work reveals that there is fine-tuned interlinked regulation between neighboring genetic elements, spanning multiple genes that determines where transcripts start and stop. The ability to predict these interactions can inform key 'design principles' for genome construction; i.e. where are [genes](#) best located and how should they be positioned relative to each other. These insights advance tools for engineering transcripts without

changing the sequence itself but by modulating neighboring gene expression." Their work adds to a growing repertoire of design principles that can be leveraged to realize a grand vision in synthetic biology: designing and building a genome from scratch.

**More information:** Aaron N. Brooks et al, Transcriptional neighborhoods regulate transcript isoform lengths and expression levels, *Science* (2022). DOI: [10.1126/science.abg0162](https://doi.org/10.1126/science.abg0162).  
[www.science.org/doi/10.1126/science.abg0162](https://www.science.org/doi/10.1126/science.abg0162)

Provided by European Molecular Biology Laboratory

Citation: Understanding genomes, piece by piece (2022, March 3) retrieved 26 June 2024 from <https://phys.org/news/2022-03-genomes-piece.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.