to quantify how the features of a design were linked to the observed measurements.

What did this enormous endeavor reveal? First, although codon choice has long been considered crucial for optimizing protein synthesis in *E. coli*, Cambray *et al.*[1] found that RNA secondary structure near the start codon is the design factor that has by far the largest effect on translational efficiency, thus underlining the importance of translation initiation. Other features, such as nucleotide, codon, and amino acid composition, have only a minor influence compared with RNA secondary structure. Codon choice—as measured by the codon adaptation index[9]—has a more pronounced effect on translation only when the translation initiation rate of all constructs is increased, thus offering some comfort for companies that tout codon-optimization tools.

The analysis identified six archetypal scenarios for translational efficiency, depending on the gene sequence. These scenarios include trade-offs between secondary structures affecting the initiation rate and codon usage influencing elongation, with an ideal balance occurring when ribosomes load onto a transcript as quickly as they progress along the mRNA during translation. In contrast, the worst-translated constructs encode multiple strong and noncompeting RNA structures that block ribosomes during translation and inhibit mRNA degradation, thereby resulting in greater accumulation of these transcripts. This accumulation sequesters a substantial portion of the cell's ribosomes in an unproductive state and thus strongly affects cell growth.

Despite the scale and detail of this study, however, the recipe for efficient translation remains far from perfected. Only around half of the measured differences in protein synthesis were accounted for by the design parameters that the authors considered. Many other factors and mechanisms therefore remain to be discovered and understood.

The lack of a complete explanation highlights a major challenge in trying to understand sequence-to-function links in biology. The combinatorics of most genetic sequences far exceeds the number of data points that can be collected. Although DoE can help maximize what is learned, it relies on an understanding of factors that have been deemed important a priori. This problem is amplified as research moves toward more complex biological systems and processes—such as protein folding, metabolic engineering, and bioprocess design—whose features must be carefully tuned in unison with the others.

Does developing a full understanding of the biological processes really matter for most applications? Unlike those in nature, the demands in industry are often highly constrained. Cells can be kept under almost constant conditions in a carefully controlled bioreactor, or a protein product may have a fixed amino acid sequence that allows for only a relatively small set of synonymous codon changes. Constraining biology, both genetically and environmentally, can help decrease the number of factors that must be considered and simplify the search for designs that produce the desired outputs.

Cambray *et al.*[1] show how far the field has come in the synthesis and testing of genetic systems *en masse*. As capabilities continue to grow, a point will inevitably be reached in which the constrained genetic-design spaces of some applications are sufficiently sampled for machine learning to become effective (**Fig. 1**). The use of machine learning would decrease the need to know which factors matter up front and allow sequence–function relationships to be deciphered from data alone[10]. The combination of DoE-based DNA-library synthesis, high-throughput measurement approaches, and modern machine-learning tools such as deep learning offers exciting prospects for those wishing to disentangle multifactorial research questions in biology and biotechnology.

COMPETING INTERESTS
The authors declare no competing interests.

1. Cambray, G. *et al. Nat. Biotechnol.* **36**, 1005–1015 (2018).
2. Gingold, H. & Pilpel, Y. *Mol. Syst. Biol.* **7**, 481 (2011).
3. Kudla, G., Murray, A.W., Tollervey, D. & Plotkin, J.B. *Science* **324**, 255–258 (2009).
4. Welch, M. *et al. PLoS One* **4**, e7002 (2009).
5. Kosuri, S. *et al. Proc. Natl. Acad. Sci. USA* **110**, 14024–14029 (2013).
6. Ingolia, N.T., Ghaemmaghami, S., Newman, J.R. & Weissman, J.S. *Science* **324**, 218–223 (2009).
7. Charneski, C.A. & Hurst, L.D. *PLoS Biol.* **11**, e1001508 (2013).
8. Gorochowski, T.E., *et al. Nucleic Acids Res.* **43**, 3022–3032 (2015).
9. Sharp, P.M. & Li, W.H. *Nucleic Acids Res.* **15**, 1281–1295 (1987).
10. Zhang, J.X. *et al. Nat. Chem.* **10**, 91–98 (2018).

Research Highlights

Papers from the literature selected by the Nature Biotechnology editors. (Follow us on Twitter, @NatureBiotech #nbtHighlight)

A homing system targets therapeutic T cells to brain cancer
Samaha, H. *et al. Nature* 10.1038/s41586-018-0499-y (2018)

RNA velocity of single cells
Manno, G. *et al. Nature* **560**, 494–498 (2018)

The chimeric TAC receptor co-opts the T cell receptor yielding robust anti-tumor activity without toxicity
Helsen, C.W. *et al. Nat. Commun.* **9**, 3049 (2018)

Post-antibiotic gut mucosal microbiome reconstitution is impaired by probiotics and improved by autologous FMT
Suez, J. *et al. Cell* **174**, 1406–1423 (2018)

Personalized gut mucosal colonization resistance to empiric probiotics is associated with unique host and microbiome features
Zmora, N. *et al. Cell* **174**, 1388–1405 (2018)