Gene regulation

https://doi.org/10.1038/s41576-025-00847-w

# Understanding the logic and grammar of *cis*-regulatory elements

*Cis*-regulatory elements (CREs), such as promoters and enhancers, possess complex architectures of transcription factors that determine specific expression patterns, levels and responses to diverse cellular or environmental cues. Understanding the rules that drive such specificity has been the holy grail of molecular biology for more than four decades. This knowledge could be used to determine disease mechanisms and design synthetic CREs for therapeutic or bioengineering purposes.

Thousands of studies from the 1980s, 1990s and 2000s used reporter assays to test the activities of individual CREs and how they are affected by selected mutations of putative transcription factor binding sites. This body of work showed that predicting CRE activity was more challenging than anticipated, suggesting that a thorough mechanistic understanding would require evaluating thousands of CRE sequences. In 2012, two seminal studies by Melnikov et al. and Patwardhan et al. tackled this grand challenge by leveraging two key advances at the time: the synthesis of large, complex oligonucleotide pools and measurement of the activity of thousands of CRE variants in parallel using high-throughput sequencing. While presenting these papers at a journal club during my postdoc, I clearly recall the sense that they would transform how CREs are studied.

Melnikov et al. developed massively parallel reporter assays (MPRAs) that measure the transcriptional activity of thousands of CRE variants by pairing CRE sequences with multiple barcodes in the 3′ untranslated region of a reporter gene, and then determining relative barcode read counts between the mRNA and DNA fractions. In this study, the authors performed saturation mutagenesis and tested thousands of mutation combinations, multi-nucleotide substitutions, and insertions to determine their impact on the activity of two inducible CREs. One major finding was that many single nucleotide mutations, even those outside of known transcription factor binding sites, can contribute to CRE activity. The authors also found some mutations that increased activity either by increasing the binding affinity of activating transcription factors or by perturbing cryptic repressor sites.

> **"I clearly recall the sense that [these papers] would transform how [*cis*-regulatory elements] are studied"**

Using a complex library with multiple mutations, Melnikov et al. generated models to quantify the dependency between enhancer activity and specific nucleotide mutations. These predictions enabled the identification of a series of substitutions that increase activity. However, the authors also showed that nucleotide mutations that increase activity in induced conditions often greatly increased basal activity, thereby reducing inducibility, which could be detrimental in biological and synthetic biology contexts. Although the models generated were relatively simple and specific to each CRE, they proved useful to design CREs with specific inducibility features. Furthermore, these pioneering models paved the way for more complex machine learning methods developed in recent years that predict activity from DNA sequence alone.

Patwardhan et al. developed a similar approach termed massively parallel functional dissection (MPFD), in which CREs are assembled using overlapping oligonucleotides with a low level of degeneracy. This method led to a more complex CRE library than those tested by Melnikov et al., and enabled testing of single nucleotide mutations as well as all possible pairs of mutations. Another major difference between MPRAs and MPFD is that libraries were evaluated in vivo in mouse livers for MPFD rather than cell lines for MPRAs. Several findings derived from this MPFD study yielded similar conclusions to those obtained by MPRAs, including that many mutations affect activity, although with moderate or mild effects, and that mutations generally have additive effects. This finding suggested that many CREs are robust to mutations and may function through relatively independent transcription factor recruitment with few epistatic interactions.

These papers were visionary and presented many ideas and future directions that have guided research in the field for more than a decade, including the use of these approaches to characterize other types of CREs (for example, core promoters, silencers and insulators), study non-coding disease variants, develop machine learning methods that predict activity from sequence, and design synthetic CREs with user-selected features. These studies illustrate how technological innovation, combined with clever design strategies, can lay the foundation for many new discoveries across fields such as disease genetics, functional genomics, evolutionary biology and bioengineering.

**Juan I. Fuxman Bass**
Biology Department, Boston University, Boston, MA, USA.
e-mail: fuxman@bu.edu

**Competing interests**
The author declares no competing interests.

**Original articles:** Melnikov, A. et al. Systematic dissection and characterization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat. Biotechnol.* **30**, 271–277 (2012); Patwardhan, R. P. et al. Massively parallel functional dissection of mammalian enhancers in vivo. *Nat. Biotechnol.* **30**, 265–270 (2012)