DECODING THE MECHANISTIC IMPACT OF GENETIC VARIATION ON REGULATORY SEQUENCES WITH DEEP LEARNING

Evan Seitz, David McCandlish, Justin Kinney, Peter Koo

Cold Spring Harbor Laboratory, Simons Center for Quantitative Biology, Cold Spring Harbor, NY

Deciphering how DNA sequences encode cis-regulatory mechanisms is a central challenge in biology. Cis-regulatory elements integrate signals—such as specific transcription factor binding sites combined with broader sequence context—to orchestrate gene expression. Although deep learning has advanced our ability to predict regulatory activity from DNA, these models have yet to systematically reveal how genetic variation reshapes the underlying regulatory mechanisms. Here, we introduce SEAM, a computational framework that uses deep learning and explainable AI to uncover how small changes in DNA can reconfigure cis-regulatory logic. SEAM generates sets of variant sequences and applies model interpretation methods to pinpoint the key nucleotide positions that together form a regulatory "mechanism"—a composite signature reflecting both transcription factor binding sites and their surrounding features. By mapping sequences into a learned "mechanism space" (a data-driven representation of regulatory signatures) and clustering those with similar profiles, SEAM reveals how individual mutations remodel regulatory DNA and drive functional diversity. Applied to human and fly regulatory elements, SEAM not only highlights the remarkable evolvability of cis-regulatory sequences but also disentangles the specific contributions of transcription factor binding from broader sequence context. In doing so, it offers critical insights into the regulatory grammar underlying gene expression and provides a robust framework for guiding the rational design of synthetic DNA with tailored functions.