https://doi.org/10.1038/s41576-025-00882-7

Cis-regulatory elements at cellular resolution

Yi Xiang See, Tim Stuart & Jay W. Shin

Recent advances in single-cell profiling technologies now enable routine and scalable measurements of *cis*-regulatory element activity across diverse cell types, disease states and genetic backgrounds. It is time to coordinate a global effort to systematically map *cis*-regulatory element function at single-cell resolution.

Cis-regulatory elements (CREs), such as promoters and enhancers, drive spatiotemporal gene expression patterns that are essential for development and organismal function and, when disrupted, can lead to human disease. Large-scale consortia efforts such as ENCODE (The Encyclopedia of DNA Elements) and FANTOM (Functional Annotation of the Mammalian Genome) have had a pivotal role in cataloguing candidate CREs, but the data collected in these efforts are primarily derived from bulk tissues or immortalized cell lines, which obscure cellular heterogeneity and often fail to recapitulate the regulatory landscape of primary cells[1]. Fortunately, single-cell technologies can now measure several important functional genomics outputs, include gene expression (single-cell RNA sequencing), chromatin accessibility (single-cell ATAC-seq) and transcription start site location (single-cell CAGE)[2,3]. These methods have revealed substantial cell type-specific CRE activity that changes dynamically in development, disease and responses to the environment. For example, a recent single-cell chromatin accessibility atlas profiled 222 fetal and adult human cell types and identified more than 500,000 previously unreported putative CREs[4], underscoring the vast regulatory diversity that single-cell approaches can resolve.

The importance of measuring and cataloguing CREs at single-cell resolution is clear, but taking full advantage of the advances in single-cell technologies will entail collecting single-cell datasets from diverse ancestries, functionally characterizing CRE variants at single-cell resolution and, ultimately, building a comprehensive single-cell atlas of the regulatory genome.

Representation across diverse ancestries

Despite the advances in single-cell technologies that are improving data resolution, representation across diverse ancestries in these data remains limited. Most datasets for characterizing CREs are heavily biased towards individuals of European ancestry, which limits the detection of CRE variants that may influence gene regulation in other populations. Furthermore, reliance on a single reference genome sequence introduces alignment bias and reduces sensitivity to structural variants and polymorphisms that may alter CRE activity. To overcome these limitations, the field is increasingly turning to the Human Pangenome Reference — a graph-based, multi-genome representation that captures structural and sequence variation across diverse populations. Aligning single-cell regulatory data to such references will enhance variant detection and improve the equity and accuracy of future functional maps[5].

Functional characterization of CRE variants

Studying the functional consequences of variation in CREs has typically involved population-scale studies or high-throughput functional assays such as massively parallel reporter assays. However, both approaches have limitations: population-scale studies are typically limited to common variants and are confounded by linkage disequilibrium and ancestry bias[6], and massively parallel reporter assays are often conducted in artificial episomal contexts or immortalized cell lines that lack endogenous genomic context. To address these limitations, new experimental perturbation tools can be combined with single-cell technologies to functionally characterize any CRE variant across heterogeneous and physiologically relevant cell types[7]. Such perturbation tools include CRISPR-based editing of CREs in their native genomic loci, as well as emerging technologies such as programmable recombinases or transposases[8,9], random genome shuffling and large-scale genome writing for more complex sequence modifications at scale. The combination of these tools with single-cell profiling technologies should be embraced both to map natural cell states and for high-throughput screening of DNA sequence variants. Such datasets would provide the basis for a deeply integrated functional map of the human genome, which should be a major focus for the field.

The combination of experimental perturbation with single-cell profiling is a powerful approach to constructing a comprehensive functional atlas of CRE activity and variant effects, but doing so will require several key components. First, these assays must use diverse cellular models, including primary cells, organoids, stem cell-derived systems and teratomas, to reflect a wide range of human cell states. Second, cross-species comparisons (supported by single-cell cross-species mapping) will be important in cases in which human cell states are experimentally intractable, as model organisms may provide insight through homologous regulatory mechanisms. Third, further technological developments will help to functionally characterize variants beyond single nucleotide polymorphisms, including complex rearrangements to regulatory architectures. Lastly, computational models and artificial intelligence will be needed to generalize regulatory logic and predict variant effects across unseen sequences, cell states and conditions. Experimental perturbation approaches that incorporate these components will enable us to systematically perturb the human genome sequence, observe the resulting effects on cell states, and perform these screens across a broad array of models that reflect diverse human cell types.

Towards a comprehensive single-cell atlas

Past and present consortia have made important progress in cataloguing CREs, but these efforts have primarily relied on bulk tissues and

a limited set of cell types. By contrast, initiatives such as the Human Cell Atlas[10] have amassed extensive single-cell transcriptomic profiles across a broad range of biological systems yet lack systematic profiling of CREs. These complementary strengths and limitations present an opportunity to align and integrate global efforts in genetics, epigenomics, functional genomics and single-cell technologies. The time to do so is now — with recent advances in experimental platforms, data generation has become incredibly accessible and scalable.

"...with recent advances in experimental platforms, data generation has become incredibly accessible and scalable"

The next frontier of CRE biology lies in coordinated, large-scale functional annotation of CREs across diverse cell types, developmental stages, disease states and ancestries. To achieve this goal, we propose the formation of a global consortium with the following aims:

- Harmonize efforts to generate single-cell atlases by collecting paired single-cell transcriptomic and epigenomic profiles across well-characterized tissues and cell types
- Incorporate whole-genome sequencing for all samples and align data to the Human Pangenome Reference to improve sensitivity for detecting population-specific regulatory variants
- Standardize experimental protocols and metadata schemas, ensuring that datasets are interoperable and adhere to FAIR (findable, accessible, interoperable, reusable) principles
- Leverage emerging technologies such as precise genome editing and multimodal single cell assays to map variant effects in physiologically relevant contexts and native chromatin environments
- Expand to cross-species atlases to characterize conserved and divergent regulatory mechanisms in homologous cell states, particularly when human models are experimentally intractable
- Integrate data with predictive artificial intelligence frameworks, enabling the modelling of CRE function across unseen genomic sequences, cellular contexts and species

Ultimately, integrating single-cell technologies with functional genomics and population-scale sequencing will advance our understanding of gene regulation, enhance variant interpretation and support the development of precision medicine strategies tailored to diverse individuals and populations.

Yi Xiang See [1], **Tim Stuart** [1] **& Jay W. Shin** [1,2] ✉
[1]Genome Institute of Singapore (GIS), Agency for Science Technology and Research (A*STAR), Singapore, Republic of Singapore.
[2]Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Republic of Singapore.
✉e-mail: jay_shin@a-star.edu.sg

Published online: 16 September 2025

References

1. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
2. Cusanovich, D. A. et al. Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910–914 (2015).
3. Moody, J. et al. A single-cell atlas of transcribed cis-regulatory elements in the human genome. Preprint at *bioRxiv* https://doi.org/10.1101/2023.11.13.566791 (2023).
4. Zhang, K. et al. A single-cell atlas of chromatin accessibility in the human genome. *Cell* **184**, 5985–6001.e19 (2021).
5. Liao, W.-W. et al. A draft human pangenome reference. *Nature* **617**, 312–324 (2023).
6. Maurano, M. T. et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
7. Martin-Rufino, J. D. et al. Massively parallel base editing to map variant effects in human hematopoiesis. *Cell* **186**, 2456–2474.e24 (2023).
8. Durrant, M. G. et al. Bridge RNAs direct programmable recombination of target and donor DNA. *Nature* **630**, 984–993 (2024).
9. Witte, I. P. et al. Programmable gene insertion in human cells with a laboratory-evolved CRISPR-associated transposase. *Science* **388**, eadt5199 (2025).
10. Regev, A. et al. The human cell atlas. *eLife* **6**, e27041 (2017).

Acknowledgements

This work is supported by the National Research Foundation, Singapore, under its NRF Fellowship programme (T.S.; NRFF16-2024-0030) and NRF Investigatorship programme (J.W.S.; NRFI08-0022), and Genome Institute of Singapore, A*STAR.

Competing interests

The authors declare no competing interests.