Tools of the trade

SCENT defines non-coding disease mechanisms using single-cell multi-omics

Genome-wide association studies (GWAS) have mapped hundreds of thousands of genomic locations that underlie disease susceptibility. However, only rarely is it possible to pinpoint causal variants or target genes from GWAS signals. More than 90% of loci act through non-coding, regulatory mechanisms, and predicting variant effects on gene expression has been extremely challenging. The translation of GWAS loci into disease mechanisms requires human tissue-specific maps of functionally active regulatory elements that harbour causal variants, as well as the genes they target. To this end, we developed SCENT (single-cell enhancer–target gene mapping), a tool to generate cell-type-specific enhancer–gene maps from single-cell multimodal ATAC-seq and RNA-seq data[1].

SCENT is a non-parametric statistical method that associates enhancer activity (that is, peak accessibility from ATAC-seq) with gene expression from RNA-seq across single cells (Fig. 1). We reasoned that an enhancer whose expression tracks with its target gene's expression is more likely to be functionally active. SCENT uses Poisson regression followed by bootstrapping to properly model sparse, discrete and highly variable gene expression in single-cell data. The resulting enhancer–gene linkages are then used to prioritize causal variants in GWAS within the SCENT enhancers and connect them to their target genes.

SCENT directly models associations between ATAC peak accessibility and RNA counts across individual single cells in a specific cell type, rather than normalizing or aggregating them, as previous methods have done. This approach resulted in more accurate enhancer–gene linkages, as validated by expression quantitative loci, CRISPR-Flow FISH and H3K27ac HiChIP. Moreover, SCENT had greater enrichment for known disease alleles from statistical fine-mapping of GWAS than the bulk-tissue enhancer–gene maps conventionally used to annotate GWAS loci.

Given that multimodal data can be obtained from nuclear material without tissue disaggregation, SCENT is applicable to primary human samples that are difficult to disaggregate. As SCENT uses each single cell as a unit in the association model, enhancer–gene links can be defined even in a tissue from a single donor with disease-relevant cell types. This is in stark contrast to bulk-tissue methods, which require hundreds of donors for the generation of robust enhancer–gene links, as each sample is a unit in the association model. Thus, SCENT can dynamically generate disease-relevant enhancer–gene links from hard-to-obtain tissues in patients that are most useful for defining disease alleles, including tissues not typically represented in large genomics consortia.

To assess SCENT's utility in defining disease mechanisms, we applied it to a challenging locus at 4p15.2, which we originally reported in a GWAS for rheumatoid arthritis and for which causal alleles and genes were ambiguous[2]. This locus contains 21 candidate variants, each with a low probability of causality, on the basis of statistical fine mapping. Bulk-tissue methods prioritized more than 10 genes. However, in T cells obtained from inflammatory joints, SCENT enhancer–gene links nominated a single causal variant (*rs35944082*) and implicated a target gene located 235 kb away: *RBPJ*. Knockdown of *Rbpj* in mice disrupted regulatory T cells[3], consistent with a role in autoimmune disease.

From annotating non-coding alleles that cause rare diseases to somatic mutational hotspots in cancer, the functional regulatory annotation by SCENT is versatile. Excitingly, as the size of single-cell multimodal datasets continues to grow, we anticipate that we will have the statistical power to define precise cell-state-specific enhancer–gene maps by SCENT. These maps will contribute to the fundamental understanding of variant effects in specific cell states. Once genetic variation is sufficiently represented in these datasets, we will also be able to investigate whether individual genotypes affect the magnitude of regulatory effects on gene expression.

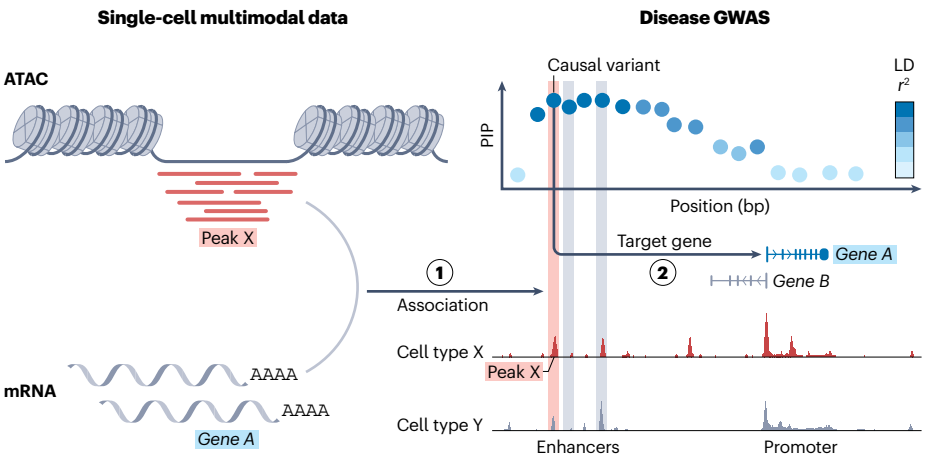

Fig. 1 | Key concept underlying SCENT. SCENT identifies (1) active *cis*-regulatory regions and (2) their target genes in a specific cell type, which can be used to define likely causal variants, genes and cell types for genome-wide association study (GWAS) loci. LD, linkage disequilibrium; PIP, posterior inclusion probability. Adapted from ref. 1.

Saori Sakaue [1,2]
[1]Center for Data Sciences and Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [2]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA.
e-mail: ssakaue@broadinstitute.org

References

1. Sakaue, S. et al. Tissue-specific enhancer-gene maps from multimodal single-cell data identify causal disease alleles. *Nat. Genet.* **56**, 615–626 (2024).
2. Ishigaki, K. et al. Multi-ancestry genome-wide association analyses identify novel genetic mechanisms in rheumatoid arthritis. *Nat. Genet.* **54**, 1640–2651 (2022).
3. Delacher, M. et al. Rbpj expression in regulatory T cells is critical for restraining T_{H2} responses. *Nat. Commun.* **10**, 1621 (2019).

Competing interests

The author declares no competing interests.