

Preview

Missing regulatory effects on complex traits: Contribution of distal variants

Sool Lee¹ and Hakhamanesh Mostafavi^{1,2,*}¹Center for Human Genetics and Genomics, New York University School of Medicine, New York, NY, USA²Department of Population Health, New York University School of Medicine, New York, NY, USA*Correspondence: Hakhamanesh.Mostafavi@nyulangone.org<https://doi.org/10.1016/j.xgen.2025.100809>

Most genetic effects on complex traits lie in non-coding regions, yet many show no regulatory activity in standard gene expression assays. In this issue of *Cell Genomics*, Arthur et al.¹ add early development-like cell types and chromatin assays, showing that distal variants missed in expression assays partly explain this discrepancy.

Genome-wide association studies (GWASs) are the standard approach for mapping genetic variants associated with human traits and diseases. By systematically analyzing genetic variation across the genome, GWASs have identified hundreds of thousands of variants associated with a broad range of phenotypes, including height, Alzheimer's disease, and type 2 diabetes. Despite this success, extracting biological insights from GWASs remains a major challenge. One reason is that most GWAS variants (around 90%) are located in non-coding regions.² These variants are presumed to act through regulatory effects, but for most, both their target genes and the contexts in which they affect gene expression remain unknown.

To address this, one approach is to test whether GWAS variants overlap with expression quantitative trait loci (eQTLs), which are variants that exhibit regulatory activity in gene expression assays. In practice, because both GWAS and eQTL loci implicate multiple variants due to linkage disequilibrium (LD), computational tools have been developed to test whether a given variant drives both GWAS and eQTL signals at a locus or whether they “colocalize.”³ However, despite the discovery of millions of eQTLs across various tissues and cell types, their overlap with GWAS variants is relatively small, leaving most GWAS loci unexplained.⁴ This has given rise to an ongoing puzzle, recently referred to as the “missing regulation” problem.⁵

There are two dominant explanations proposed for the problem of missing regulation: (1) missing GWAS variants

are eQTLs that are context specific—such as being transiently active during development or induced upon stimulation—and are therefore overlooked by standard eQTL assays, which typically analyze postmortem bulk tissues,⁶ and (2) eQTL mapping is systematically biased away from trait-relevant genes, even when the right tissues and contexts are sampled, partly due to the effect of natural selection on human phenotypes.⁷ Moreover, variants with small regulatory effects, such as distal variants, are difficult to detect as eQTLs at current sample sizes but can still appear as GWAS hits if they regulate high-impact genes.⁷

Arthur et al. set out to test these scenarios by expanding QTL mapping in two ways. First, they studied early development-like cell types, specifically induced pluripotent stem cells (iPSCs), iPSC-derived cardiovascular progenitor cells (CVPCs), and iPSC-derived pancreatic progenitor cells (PPCs), to assess how much of the missing regulation is due to unexplored contexts. Second, they complemented eQTLs with genetic effects on chromatin accessibility (caQTLs) and histone acetylation (haQTLs), a marker of active promoters and enhancers. While gene-proximal eQTLs typically have larger effects and are easier to detect, chromatin QTL discovery is likely independent of gene distance and may therefore improve colocalization detection at distal regulatory elements. Consistent with this expectation, chromatin QTLs showed weaker enrichment at promoters. However, it is important to note that if these assays had infinite statistical power, they should

converge on the same set of variants, as they probe molecular intermediates along the same regulatory pathway (Figure 1A).

Across all development-like cell types and QTLs, Arthur et al. found colocalization for ~10% of GWAS loci across 15 traits (540 out of 5,192). This is a relatively small number, likely due to the limited sample sizes they used for QTL mapping, on the order of a few hundred individuals. Nevertheless, colocalization patterns in these data provide insights into the missing regulation problem. Notably, incorporating chromatin QTLs led to a 2.3-fold increase in colocalization rates compared to eQTLs alone. Interestingly, GWAS-ca/haQTL colocalized variants were, on average, more distal than GWAS-eQTL colocalized variants. Furthermore, GWAS variants without any colocalization were farther from genes than colocalized GWAS variants. These observations support the model that some of the GWAS-eQTL gap arises from small regulatory effects of distal GWAS variants that are difficult to detect in gene expression assays.⁷ These variants may be better captured by chromatin QTLs (Figure 1B).

Strikingly, the majority of eQTLs identified in the early development-like samples (i.e., “EDev-like”) were also present in adult tissues; only ~3% (2,299 out of 74,494 eQTL-gene pairs) were specific to EDev samples. These EDev-specific eQTLs exhibited smaller effects than shared eQTLs did, though this could arise if they tend to be more gene distal. More interestingly, EDev-specific eQTLs contributed minimally to colocalized GWAS signals (13 of 540). A similar



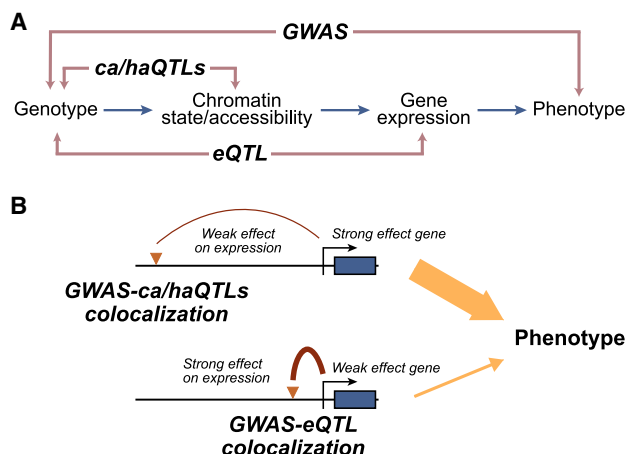


Figure 1. Complementary methodologies to study the regulatory function of genetic associations

(A) Genetic variation can influence chromatin accessibility and state, gene expression, and ultimately phenotypic traits through a regulatory cascade. GWASs identify genotype-phenotype associations, while QTL mapping detects genetic associations with regulatory intermediates, including gene expression (eQTLs), chromatin accessibility (caQTLs), and histone acetylation (haQTLs).

(B) Colocalization patterns of two GWAS hits (triangles) with similar phenotypic impact. One variant has a smaller effect on expression but regulates a higher-impact gene. eQTL signals are typically stronger for gene-proximal variants, potentially missing distal variants, which may be better captured by chromatin QTLs.

pattern was observed in a previous study examining eQTLs during iPSC differentiation toward neural fates.⁸ Together, these findings suggest that broadly sampling additional developmental stages may not provide a straightforward solution to the missing regulation problem.

Lastly, the authors demonstrated the value of their multiomic QTLs in prioritizing variants within GWAS loci. Because nonfunctional variants can appear statistically associated with a trait due to LD, fine-mapping approaches have been developed to identify those most likely to have functional effects, defining them within a “credible set.” The authors show that integrating multiple lines of functional evidence—including colocalizing GWAS signals with different QTL types—can help narrow these credible sets to a smaller number of putative causal variants. Notably, most variants identified through this approach differed from those prioritized solely by statistical probability, consistent with recent evidence that the top statistical candidate is not always functional.⁹ While powerful, this strategy is limited to GWAS loci with sufficient functional data.

In conclusion, the work of Arthur et al. advances our understanding of the missing regulation problem, providing

evidence consistent with both dominant explanations previously proposed. While it is essential to sample trait-relevant cell types and contexts, genetic effects on complex traits are possibly distributed across a broad range of regulatory contexts. As a result, studying any single context may not substantially increase overall colocalization, even though each newly explored context will likely contribute additional colocalized signals. Moreover, even in the most relevant cellular context, eQTL assays may be underpowered to detect distal regulatory variants. Chromatin QTLs may be particularly valuable in capturing these variants and could be complemented by computational approaches that predict chromatin architecture and regulatory landscapes from DNA sequence (e.g., Pampari et al.¹⁰). Ultimately, there seems to be no single solution to the missing regulation problem, and addressing it will likely require a multi-pronged approach.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Arthur, T.D., Nguyen, J.P., Henson, B.A., D’Antonio-Chronowska, A., Jauregui, J., Silva, N.,

Panopoulos, A.D., Izpisua Belmonte, J.C., D’Antonio, M., McVicker, G., et al. (2025). Multiomic QTL mapping reveals phenotypic complexity of GWAS loci and prioritizes putative causal variants. *Cell Genom.* 5, 100775. <https://doi.org/10.1016/j.xgen.2025.100775>.

2. Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. *Science* 337, 1190–1195. <https://doi.org/10.1126/science.1222794>.
3. Hormozdizari, F., van de Bunt, M., Segrè, A.V., Li, X., Joo, J.W.J., Bilow, M., Sul, J.H., San-karaman, S., Pasaniuc, B., and Eskin, E. (2016). Colocalization of GWAS and eQTL signals detects target genes. *Am. J. Hum. Genet.* 99, 1245–1260. <https://doi.org/10.1016/j.ajhg.2016.10.003>.
4. GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 369, 1318–1330. <https://doi.org/10.1126/science.aaz1776>.
5. Connally, N.J., Nazeen, S., Lee, D., Shi, H., Stamatoiyannopoulos, J., Chun, S., Cotsapas, C., Cassa, C.A., and Sunyaev, S.R. (2022). The missing link between genetic association and regulatory function. *Elife* 11, e74970. <https://doi.org/10.7554/elife.74970>.
6. Umans, B.D., Battle, A., and Gilad, Y. (2021). Where are the disease-associated eQTLs? *Trends Genet.* 37, 109–124. <https://doi.org/10.1016/j.tig.2020.08.009>.
7. Mostafavi, H., Spence, J.P., Naqvi, S., and Pritchard, J.K. (2023). Systematic differences in discovery of genetic effects on gene expression and complex traits. *Nat. Genet.* 55, 1866–1875. <https://doi.org/10.1038/s41588-023-01529-1>.
8. Jerber, J., Seaton, D.D., Cuomo, A.S.E., Kumasaka, N., Haldane, J., Steer, J., Patel, M., Pearce, D., Andersson, M., Bonder, M.J., et al. (2021). Population-scale single-cell RNA-seq profiling across dopaminergic neuron differentiation. *Nat. Genet.* 53, 304–312. <https://doi.org/10.1038/s41588-021-00801-6>.
9. Lee, S., McAfee, J.C., Lee, J., Gomez, A., Ledford, A.T., Clarke, D., Min, H., Gerstein, M.B., Boyle, A.P., Sullivan, P.F., et al. (2025). Massively parallel reporter assay investigates shared genetic variants of eight psychiatric disorders. *Cell* 188. <https://doi.org/10.1016/j.cell.2024.12.022>.
10. Pampari, A., Shcherbina, A., Kvon, E.Z., Kosicki, M., Nair, S., Kundu, S., Kathiria, A.S., Risca, V.I., Kuningas, K., Alasoo, K., et al. (2025). ChromBPN: bias factorized, base-resolution deep learning models of chromatin accessibility reveal cis-regulatory sequence syntax, transcription factor footprints and regulatory variants. Preprint at bioRxiv. <https://doi.org/10.1101/2024.12.25.630221>.