Cell Genomics

CellPress
OPEN ACCESS

Preview

Partitioning genetic pleiotropy within tissue-specific regulatory patterns

Prasun Panja,[1,2,3] Diptanil Biswas,[1,2,3] and Samsiddhi Bhattacharjee[1,2,*]
[1]Regional Centre for Biotechnology, Faridabad, India
[2]BRIC-National Institute of Biomedical Genomics, Kalyani, India
[3]These authors contributed equally
*Correspondence: sb1@nibmg.ac.in
https://doi.org/10.1016/j.xgen.2026.101212

GWAS-associated variants of related complex traits indicate the presence of pleiotropic clusters often related to tissue-specific epigenomic profiles. The authors introduce J-PEP, a novel framework that enables tissue-aware interpretation of complex trait genetic architecture and PEPA, a cross-modal validation metric. J-PEP factorizes trait and tissue data matrices jointly to identify soft clusters of SNPs that capture both pleiotropic and shared epigenomic patterns in tissues, improving both clustering efficiency and mechanistic interpretation.

With the advent of genome-wide association studies (GWASs), thousands of genetic loci have been implicated across a wide range of complex traits and diseases, creating an urgent need to translate statistical associations into biological understanding. Many loci are associated with multiple phenotypes, termed as pleiotropic variants, suggesting the presence of shared molecular pathways rather than isolated trait-specific mechanisms. In parallel, rapidly expanding epigenomic maps have begun to illuminate the regulatory landscapes in which these variants function, providing crucial tissue- and cell-type-specific context. A key emerging question is therefore how shared genetic signals can be systematically translated into biologically meaningful processes acting in disease-relevant tissues and cell types. Integrative frameworks that partition the pleiotropic effect and epigenomic effect simultaneously, leveraging cross-trait genetic architecture and tissue-specific regulatory information, offer a path forward, revealing latent biological axes underlying shared genetic architecture.

To identify common genetic susceptibilities across traits, several methods have employed pleiotropic clustering, which groups variants according to shared multi-trait association patterns.[1–3] Statistically, these approaches are often formulated as multi-trait matrix factorization models that decompose genetic effect matrices into a set of latent components. While such frameworks have provided important insights into shared genetic architecture, they typically lack explicit regulatory context. As a result, clusters with similar trait signatures may arise from distinct tissues or cellular environments, limiting biological interpretability. Epigenomic enrichment analyses have partially addressed this challenge,[3,4] but integrating regulatory information directly into the clustering framework remains a key unmet need. In their recent work, Kerner et al.[5] introduce J-PEP, a joint statistical model that integrates genetic pleiotropy with epigenomic context, shifting GWAS interpretation from trait-level correlations toward mechanistic partitioning of disease-associated loci (see Figure 1).

The joint pleiotropic and epigenomic partitioning (J-PEP) framework is an integrative approach designed to uncover shared biological programs underlying complex trait genetics. J-PEP operates by fixing a focal trait and incorporating a set of genetically associated auxiliary traits, leveraging two complementary representations of genetic variation. First, it constructs an SNP-by-trait matrix V_{trait}, obtained as products of posterior inclusion probabilities (PIPs) across the focal and auxiliary traits, thereby capturing cross-phenotype pleiotropic signals. Second, it incorporates an SNP-by-tissue matrix V_{tissue}, derived from epigenomic annotations that reflect regulatory activity across tissues and biological contexts. Using an extended Bayesian non-negative matrix factorization framework,[6] J-PEP jointly decomposes these matrices into a shared SNP-to-cluster membership matrix together with trait- and tissue-specific loading profiles. The factorization is obtained by minimizing an objective function comprising three elements: reconstruction errors for V_{trait} and V_{tissue} based on their model-implied factorization, and a penalty term based on the statistical significance of positive correlation between the two modalities. To enhance interpretability and stability, the model imposes a sparsity constraint on tissue assignments, such that each tissue contributes to, at most, a single SNP cluster. Through this joint modeling strategy, J-PEP identifies latent disease components that are both genetically shared across traits and grounded in tissue-specific regulatory biology.

The performance of clustering frameworks is traditionally evaluated using reconstruction error or cross-validation-based prediction accuracy. While widely used, such metrics can be sensitive to the chosen number of clusters and may introduce overfitting biases. To address this limitation, the authors introduce a new metric, pleiotropic and epigenomic prediction accuracy (PEPA), which explicitly evaluates consistency between pleiotropic and epigenomic modalities. PEPA combines pleiotropic prediction accuracy (PPA) and epigenomic prediction accuracy (EPA) using a geometric mean,

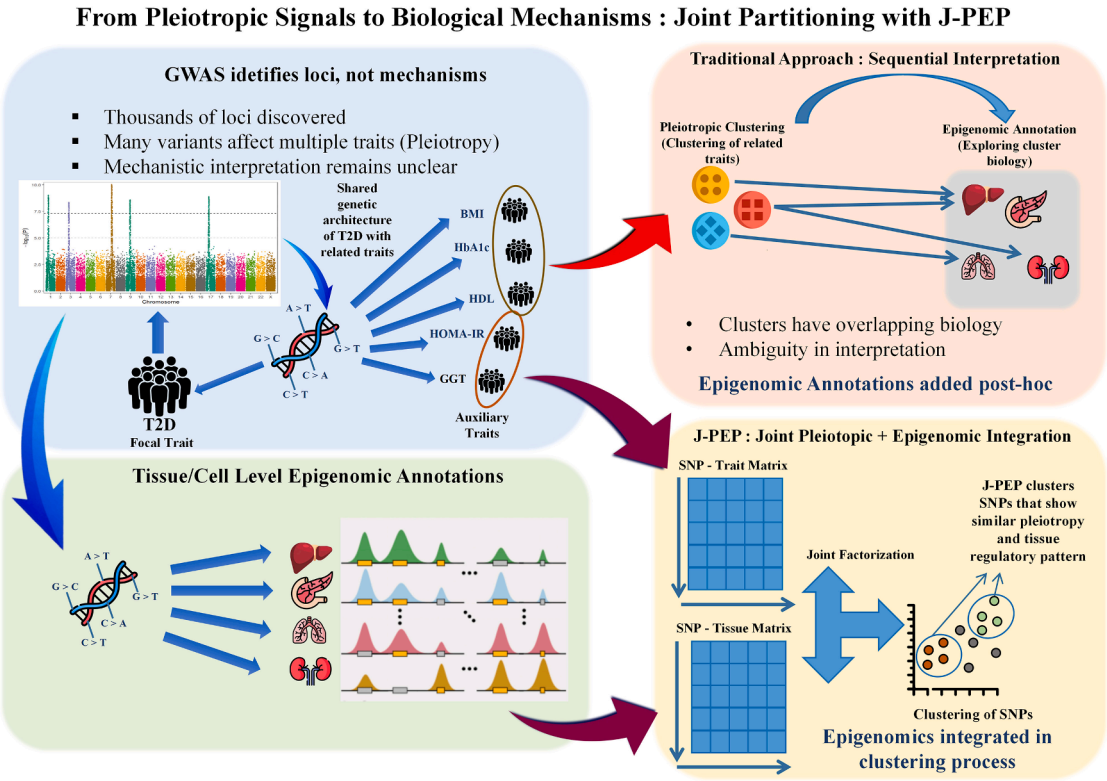

Figure 1. Traditional SNP-clustering approaches interpret pleiotropic and epigenomic signals sequentially
J-PEP jointly models SNP-trait and SNP-tissue architectures to produce tissue-informed, biologically coherent variant clusters.

thereby favoring clusters supported by both trait associations and tissue-specific regulatory signals while penalizing models dominated by a single modality. Importantly, the metric adopts an off-chromosome cross-validation framework in which one modality is used to predict the other and performance is quantified through correlations between predicted and observed matrices rather than simple reconstruction error. By emphasizing cross-modal prediction rather than within-modal fit, PEPA reduces sensitivity to cluster number and mitigates overfitting. Overall, this framework provides a statistically principled and biologically interpretable benchmark for evaluating integrative clustering approaches and highlights the value of jointly modeling genetic pleiotropy and epigenomic context.

Across various simulation settings and real data analyses, J-PEP consistently achieved higher cross-modal prediction accuracy judged by the PEPA metric compared to unimodal pleiotropic or epigenomic partitioning approaches such as Flashier[7] and FactorGO,[1] indicating that joint modeling of modalities improves recovery of biologically coherent clusters. Notably, performance gains increased with stronger genetic signals, suggesting that shared pleiotropic and epigenomic structure becomes more detectable as statistical power grows. Applied to multiple complex traits, including type 2 diabetes (T2D) and hypertension (HTN), J-PEP recapitulated established disease mechanisms while revealing additional tissue-anchored regulatory axes. In T2D, canonical pancreatic and hepatic metabolic pathways were recovered alongside immune and epithelial tissue components, while in HTN, stromal and endocrine tissue signals emerged, illustrating how integrating regulatory context refines interpretation beyond trait-level clustering alone.

Compared with pleiotropic-only methods that emphasize shared genetic effects or epigenomic-only models that capture broad tissue enrichment, J-PEP identifies clusters supported by both genetic architecture and tissue specificity. Incorporating single-cell chromatin accessibility data further sharpened these insights by resolving bulk tissue signals into specific cellular contexts, demonstrating that pleiotropic genetic signals localize to defined cellular programs rather than reflecting broad tissue-level enrichment.

Thus, the J-PEP framework proposed by Kerner and colleagues enables partitioning of the modality-specific genetic signatures into components with shared pleiotropic and epigenomic patterns. Its scalability across traits and annotations also highlights its potential for future large-scale genomic discovery. Nonetheless, several avenues for methodological refinement remain. Reducing biases due to differential annotation depth among tissues and relaxing the strict sparsity constraint on tissue assignments may further improve robustness and applicability in diverse biological settings. Although causal inference at the level of individual variants remains an open challenge, J-PEP establishes a foundation for developing disease-mechanism-guided polygenic scores and for further

methodological developments toward broader multi-omic integrative frameworks.

ACKNOWLEDGMENTS

The authors would like to acknowledge fellowship and funding support from BRIC-National Institute of Biomedical Genomics, Kalyani, India.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Zhang, Z., Jung, J., Kim, A., Suboc, N., Gazal, S., and Mancuso, N. (2023). A scalable approach to characterize pleiotropy across thousands of human diseases and complex traits using GWAS summary statistics. Am. J. Hum. Genet. *110*, 1863–1874. https://doi.org/10.1016/j.ajhg.2023.09.015.

2. Kim, H., Westerman, K.E., Smith, K., Chiou, J., Cole, J.B., Majarian, T., Von Grotthuss, M., Kwak, S.H., Kim, J., Mercader, J.M., et al. (2023). High-throughput genetic clustering of type 2 diabetes loci reveals heterogeneous mechanistic pathways of metabolic disease. Diabetologia *66*, 495–507. https://doi.org/10.1007/s00125-022-05848-6.

3. Qi, G., Chhetri, S.B., Ray, D., Dutta, D., Battle, A., Bhattacharjee, S., and Chatterjee, N. (2024). Genome-wide large-scale multi-trait analysis characterizes global patterns of pleiotropy and unique trait-specific variants. Nat. Commun. *15*, 6985. https://doi.org/10.1038/s41467-024-51075-5.

4. Suzuki, K., Hatzikotoulas, K., Southam, L., Taylor, H.J., Yin, X., Lorenz, K.M., Mandla, R., Huerta-Chagoya, A., Melloni, G.E.M., Kanoni, S., et al. (2024). Genetic drivers of heterogeneity in type 2 diabetes pathophysiology. Nature *627*, 347–357. https://doi.org/10.1038/s41586-024-07019-6.

5. Kerner, G., Kamitaki, N., Strober, B., and Price, A.L. (2026). Mapping disease loci to biological processes via joint pleiotropic and epigenomic partitioning. Cell Genom. *6*, 101138. https://doi.org/10.1016/j.xgen.2025.101138.

6. Udler, M.S., Kim, J., Von Grotthuss, M., Bonàs-Guarch, S., Cole, J.B., Chiou, J., Christopher D Anderson on behalf of METASTROKE and the ISGC, Boehnke, M., Laakso, M., Atzmon, G., et al. (2018). Type 2 diabetes genetic loci informed by multi-trait associations point to disease mechanisms and subtypes: A soft clustering analysis. PLoS Med. *15*, e1002654. https://doi.org/10.1371/journal.pmed.1002654.

7. Willwerscheid, J., Carbonetto, P., and Stephens, M. (2025). ebnm: An R Package for Solving the Empirical Bayes Normal Means Problem Using a Variety of Prior Families. J. Stat. Softw. *114*. https://doi.org/10.18637/jss.v114.i03.