Title: Tissue-resolved multivariate epigenomic patterns enable mechanistic subtyping of complex diseases

Background: Complex diseases often exhibit heterogeneity in their biological mechanisms that are difficult to dissect through conventional genome-wide association studies (GWAS) and phenotype-down subtyping. Aggregating variant effect across cases obscures context-specific regulatory mechanisms that govern risk in patient subsets. By using machine learning to examine hundreds of cell state-specific patterns of epigenomic regulation, we developed an atlas of combinatorial epigenomic annotations and demonstrated their use for parsing variant contributions to disease subtypes.

Methods: Here, we analyse epigenomic data from 833 EpiMap biosamples spanning eight chromatin marks to construct a generalizable machine learning framework that integrates and deconvolves tissue and cell-type-specific regulation. The framework learns 720 latent regulatory signatures termed *Epigenetically Co-Modulated Patterns* (EpiCops) and perform unsupervised clustering on GWAS disease variants using these signatures. Variant clusters were further evaluated by their associations to various sub-disease phenotypes in UK Biobank using inverse variance meta-analysis, cluster partitioned polygenic risk score and time-to-events models.

Results: Using EpiCops as functional annotation yielded comparable fine mapping accuracy across 25 GWAS traits comparable to benchmark method. To test its utility on variant subtyping, we first applied EpiCops on diverse composite GWAS phenotypes. EpiCops effectively parsed disease-associated variants into their ground truth diseases and, using EpiMap metadata, accurately revealed their disease-specific cell type mechanisms in an unsupervised manner. We next applied EpiCops to a heterogenous disease phenotype, Type 2 Diabetes (T2D), where recent supervised subtyping methods offer a benchmark. By analysing 643 T2D variants based on EpiCop annotations, we reveal variant clusters with opposing associations to sub-disease phenotypes including BMI/WHR, lipids, liver, and kidney function. Furthermore, we show that EpiCop-partitioned polygenic risk scores (PRS) explain more variance than whole-genome PRS by stratifying variants across subtype-informative traits. While T2D is associated with increased heart disease, we identify variant clusters with opposing associations to cardiovascular comorbidities. Stratifying patients based on low vs high risk variants reveal an 11% difference in acute cardiac incidence between these patient subgroups. To establish mechanistic insights, we used EpiCops mapped to EpiMap metadata to determine the cell type and epigenomic regulatory basis of opposing heart disease risk profiles in T2D. These results collectively provide a new unsupervised reference for T2D subtypes.

Conclusion: Taken together, EpiCops presents a scalable, interpretable framework which learns combinatorial regulatory programs to functionally subtype complex diseases, revealing tissue specific mechanisms, identifying comorbidity signatures, and improving risk prediction when genetic burden is partitioned by context-specific mechanism.