**Gerstein lab experiencing in using single-cell data to map cell-type specific eQTLs using standard approaches:**

We have extensive experience in analyzing single-cell data to infer cell-type-specific eQTLs. In (Emani *et al.* 2024), we integrated single-nucleus, multi-omics datasets from the PsychENCODE consortium to create a uniformly processed resource comprising >2.8M nuclei from the prefrontal cortex across 388 individuals (including healthy controls, and those with schizophrenia, autism spectrum disorder, and Alzheimer’s disease). Across 28 distinct cell types, we identified >1.4M single-cell eQTLs (termed “scQTLs”), with an average of >85K cis-eQTLs per cell type and ~690 eGenes per cell type. Our scQTLs are strongly enriched in narrow regions around the transcription start sites, and many of them are cell-type-specific (though ~47% of them appear in more than one cell type). To perform our scQTLs search, we followed the same general procedure used by GTEx, including conservative filtering on the cell-type level when generating pseudobulk data. About 30% of the scQTLs overlap with bulk cis-eQTLs.

**Gerstein lab experiencing in using single-cell data to map cell-type specific eQTLs using a Bayesian framework:**

Because of the sparsity intrinsic to snRNA-seq data, particularly for the rarer cell types, we implemented a Bayesian linear mixed-effects model to identify more scQTLs in (Emani *et al.* 2024). Specifically, we quantified the relationship between genotype dosage and gene expression using a Bayesian linear mixed effects model, where the QTL effect size from each cell type (for example, the effect size *βAstro* in the cell type Astrocytes) is estimated by considering a prior distribution p(*βAstro*) ~ Normal (*Θ*, *Σ*). In turn, *Θ* and Σ also have their respective prior distributions, and our objective is to approximate the joint posterior distribution p(*βAstro, βOPC, βEndothelial, … βOligo, Θ, Σ, σ-2 | X, Y*), with *X* denoting the genotypes and Y the gene expression within a given cell type. Because the joint posterior distribution has no closed-form expression, we used Gibbs sampling to approximate this distribution. Under this scheme, the hierarchical nature of the system’s setup allows for effects to be shared between cell types via global parameters *Θ* and *Σ*.

**Gerstein lab experiencing in finding isoQTLs:**

Identifying differences in isoform expression will be highly important, particularly in the human brain. We have broad experience in identifying genes with splicing-isoform QTLs (isoQTLs). For example, we identified >2.6 million isoQTLs from short-read bulk RNA-Seq of the adult brain using a standardized GTEx pipeline (Wang *et al.* 2018). We recently identified isoQTLs in 388 short-read single-cell RNA-Seq datasets from the adult dorsolateral prefrontal cortex (Emani *et al.* 2024).

**Gerstein lab experience in allele-specific expression and allele-specific binding**

We have extensive experience in analyzing allele-specific expression (ASE) and allele-specific binding (ASB). We developed a computational pipeline AlleleSeq (Rozowsky *et al.* 2011) which was originally used for identifying and quantifying ASE and ASB in GM12878. We have applied our tool to a broad spectrum of personal and functional genomics data. We have used it in multiple large Consortium projects and publications, including ENCODE and the 1000 Genomes Project (Djebali *et al.* 2012; Gerstein *et al.* 2012; Khurana *et al.* 2013). Our results were made available as an online resource, AlleleDB (alleledb.gersteinlab.org). Furthermore, using the extensive Roadmap dataset, we constructed a high-resolution map that reveals allelic imbalances in DNA methylation, histone marks, and transcription across 71 epigenomes from 36 distinct cell and tissue types from 13 donors (Onuchic *et al.* 2018).

As part of these and other projects, we developed our pipeline for ASE and ASB analysis. Firstly, to address the inherent variability in functional genomics readcount data, we have implemented the beta-binomial test to determine the significance of allelic imbalances. Additionally, we added supplementary filters to mitigate potential biases stemming from ambiguous mapping (Chen *et al.* 2016a). And we expanded the tool to call allele-specific genomic elements, such as genes or regulatory regions, giving rise to our current tool, AlleleSeq2 (Onuchic *et al.* 2018). Most recently, we applied AlleleSeq2 to the EN-TEx resource (Rozowsky *et al.* 2023) encompassing ~1.6K datasets from four donors (~30 tissues x 15 assays) (Figure 3).

A diagram of a dna model

Description automatically generated with medium confidence

We observed a significant increase in detected allele-specific events when aggregating genomic reads from multiple tissues (as opposed to a simple union of sample-specific analyses). We generated a catalog of >1M allele-specific loci (Rozowsky *et al.* 2023). Combining EN-TEx with existing genome annotations revealed strong associations between ASE and GWAS loci.

**Figure 3. An overview of the EN-TEx project, including ASE and transferQTL.**

## A diagram of a model Description automatically generatedGerstein lab experience with Transfer eQTLs

By combining the EN-TEx chromatin data and the GTEx eQTL catalog, we developed a model that transfers the activity of an eQTL from a given donor tissue to another tissue by considering the EN-TEx chromatin profile in the target tissue (Figure 7A). Our predictions were highly accurate across different tissue pairs. We applied this model to a set of 1.5 million blood eQTLs from a large-cohort study. Doing so enhanced the GTEx catalog with 500,000 new candidate eQTLs per tissue (Vosa *et al.* 2021). As expected, we also found that variants with observable chromatin activity (especially H3K36me3) were more likely to be transferred. We observed the opposite for variants associated with genes that are highly tissue-specific or have distant transcription-start sites.

**Figure 7: A) Previous transferQTL workflow. B) Proposed transferQTL2 workflo.**

**References:**

Chen, J., J. Rozowsky, T. R. Galeev, A. Harmanci, R. Kitchen *et al.*, 2016a A uniform survey of allele-specific binding and expression over 1000-Genomes-Project individuals. Nat Commun 7**:** 11101.

Djebali, S., C. A. Davis, A. Merkel, A. Dobin, T. Lassmann *et al.*, 2012 Landscape of transcription in human cells. Nature 489**:** 101-108.

Emani, P. S., J. J. Liu, D. Clarke, M. Jensen, J. Warrell *et al.*, 2024 Single-cell genomics and regulatory networks for 388 human brains. Science 384**:** eadi5199.

Gerstein, M. B., A. Kundaje, M. Hariharan, S. G. Landt, K. K. Yan *et al.*, 2012 Architecture of the human regulatory network derived from ENCODE data. Nature 489**:** 91-100.

Khurana, E., Y. Fu, V. Colonna, X. J. Mu, H. M. Kang *et al.*, 2013 Integrative annotation of variants from 1092 humans: application to cancer genomics. Science 342**:** 1235587.

Rozowsky, J., A. Abyzov, J. Wang, P. Alves, D. Raha *et al.*, 2011 AlleleSeq: analysis of allele-specific expression and binding in a network framework. Mol Syst Biol 7**:** 522.

Rozowsky, J., J. Gao, B. Borsari, Y. T. Yang, T. Galeev *et al.*, 2023 The EN-TEx resource of multi-tissue personal epigenomes & variant-impact models. Cell 186**:** 1493-1511 e1440.

Wang, D., S. Liu, J. Warrell, H. Won, X. Shi *et al.*, 2018 Comprehensive functional genomic resource and integrative model for the human brain. Science 362.