

Cell Genomics, Volume 4

Supplemental information

**Cell type and dynamic state govern
genetic regulation of gene expression
in heterogeneous differentiating cultures**

**Joshua M. Popp, Katherine Rhodes, Radhika Jangi, Mingyuan Li, Kenneth Barr, Karl
Tayeb, Alexis Battle, and Yoav Gilad**

Supplementary Materials for
Cell-type and dynamic state govern genetic regulation of gene expression in
heterogeneous differentiating cultures

Joshua M. Popp, Katherine Rhodes, Radhika Jangi, Mingyuan Li, Kenneth Barr, Karl Tayeb,
Alexis Battle, Yoav Gilad

Corresponding author: ajbattle@jhu.edu and gilad@uchicago.edu

The PDF file includes:

Figs. S1 to S9

Other Supplementary Materials for this manuscript include the following:

Supplementary Tables S1 to S13

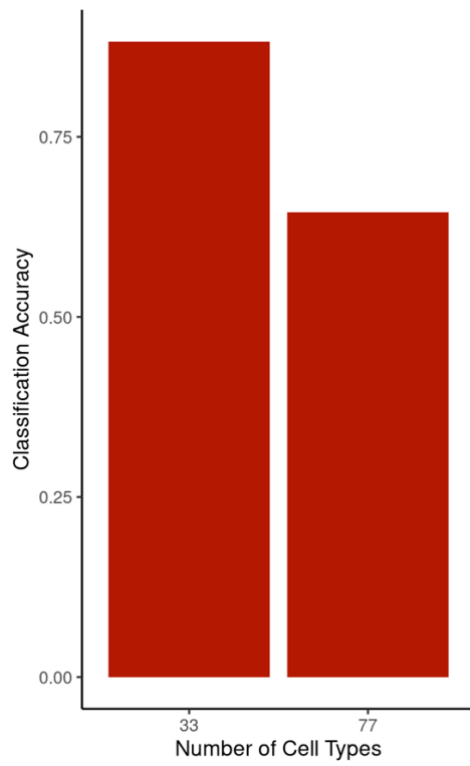


Figure S1. Classifier Comparison, Related to STAR Methods. Using original fetal cell atlas labels as ground truth, comparison of classification accuracy based on all 77 cell-types or a refined set of 33 cell-types.

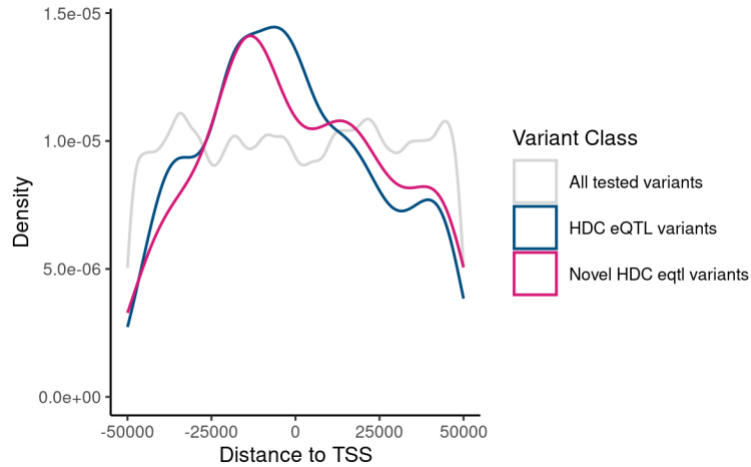


Figure S2. Genomic distribution of HDC eQTLs, Related to Figure 2. Genomic distribution of all variants tested for eQTLs of tissue development genes (all tested variants, gray), the subset of these which had significant eQTL variants (HDC eQTL variants, dark blue), and the subset of these eQTL variants which were not detected anywhere in GTEx (novel eQTL variants, pink).

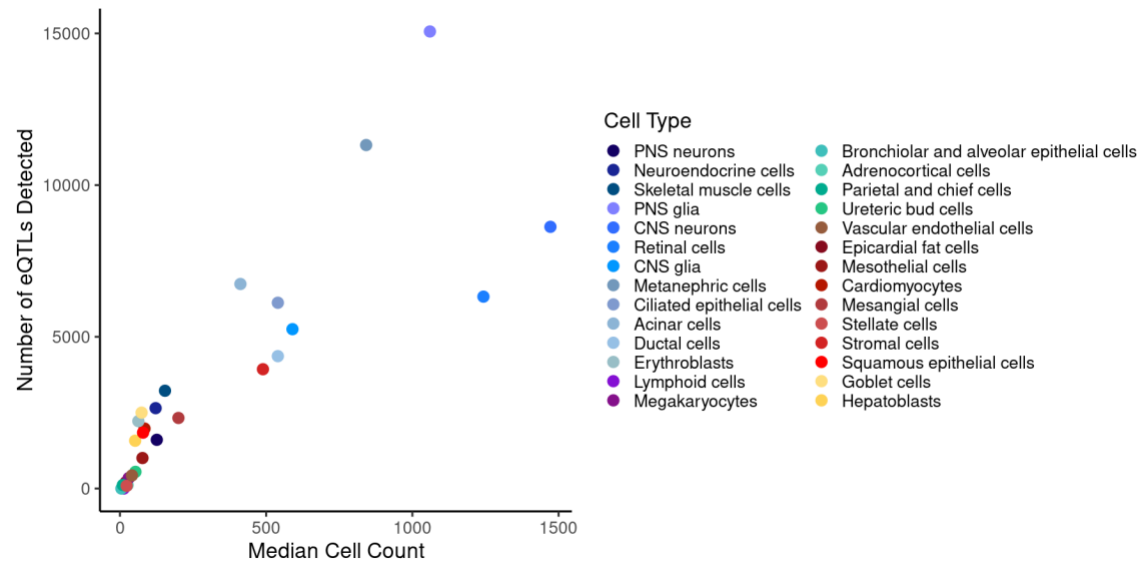


Figure S3. eQTL Detection per Cell Type, Related to Figure 2. Number of eQTLs detected per cell-type, compared to the median cell count across all samples.

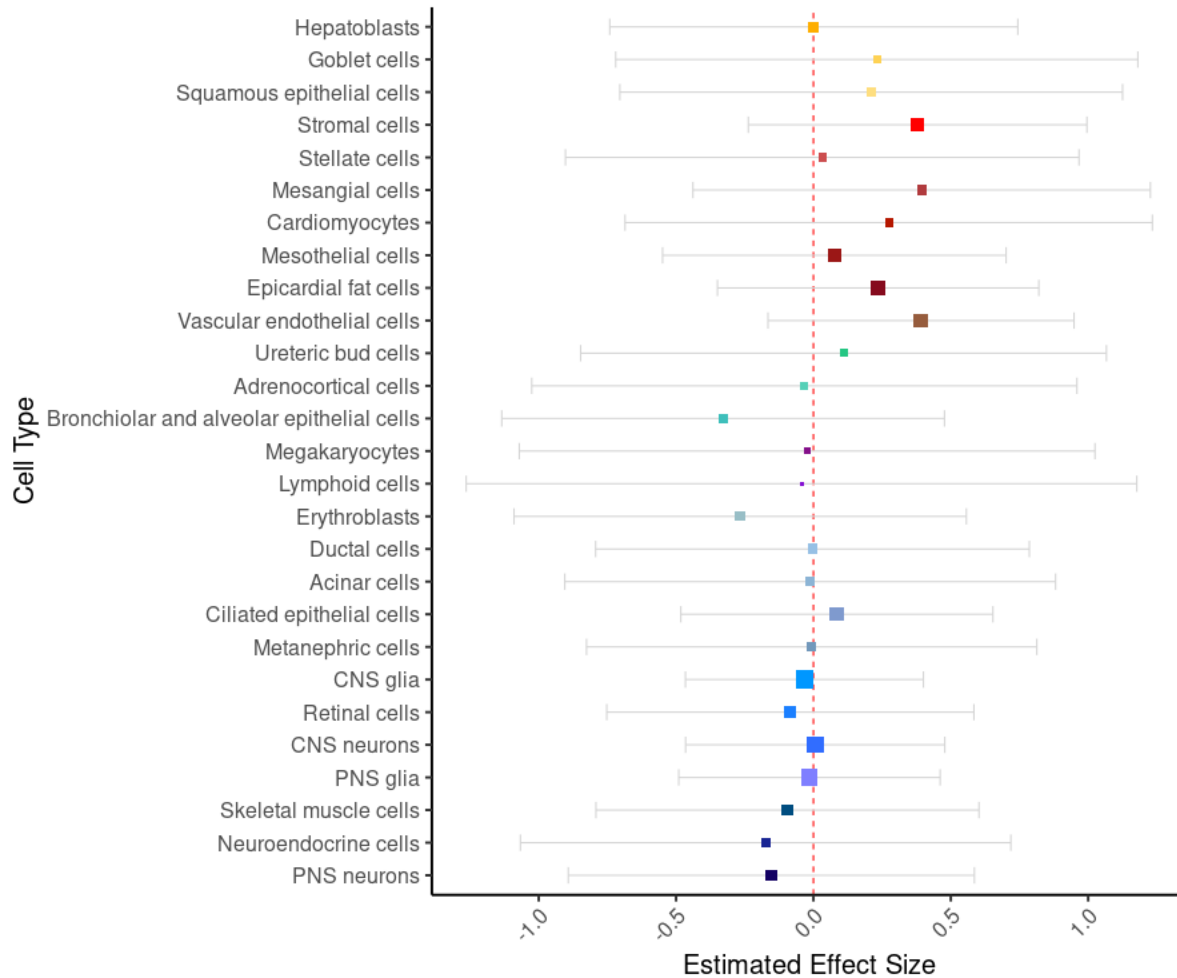


Figure S4. eQTL effect estimates without MASH, Related to Figure 2. eQTL effect estimates for the gene *SH3PXD2B* at rs10042482, based on eQTL calling in each cell type in isolation (no meta-analysis performed). Boxes are centered at the estimated effect size in the given cell type, error bars show +/- standard deviation, box size indicates precision (1/squared standard deviation).

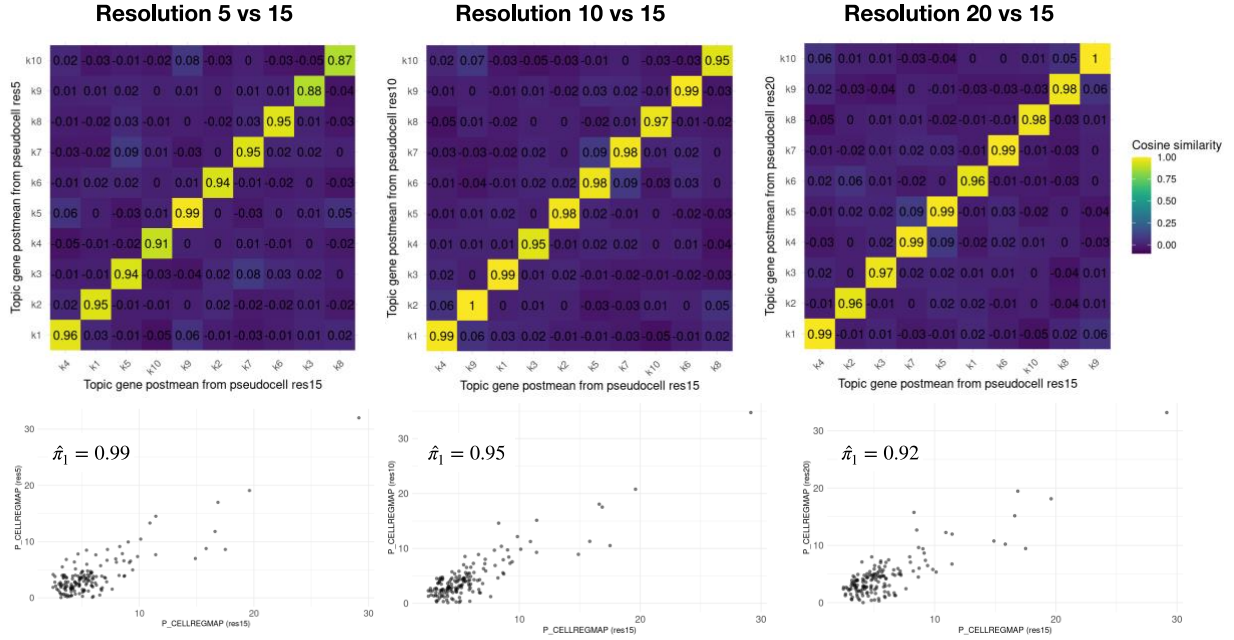


Figure S5. Topic Modeling Robustness to Hyperparameter Selection, Related to STAR Methods. (Top) Cosine similarity between topic DE vectors for a 10-topic model at varying pseudocell clustering resolutions (Leiden resolution parameter = 5, 10, 20 versus the original resolution 15). Topic DE vectors estimated using grade of membership differential expression analysis (see Methods). (Bottom) Comparison of $-\log_{10}$ P-values of reported significant topic eQTLs discovered using Leiden clustering resolution of 15 (x-axis), to the $-\log_{10}$ P-values obtained when aggregating at various resolutions using a 10-topic model. Inset shows estimated replication rate, $\hat{\pi}_1$.

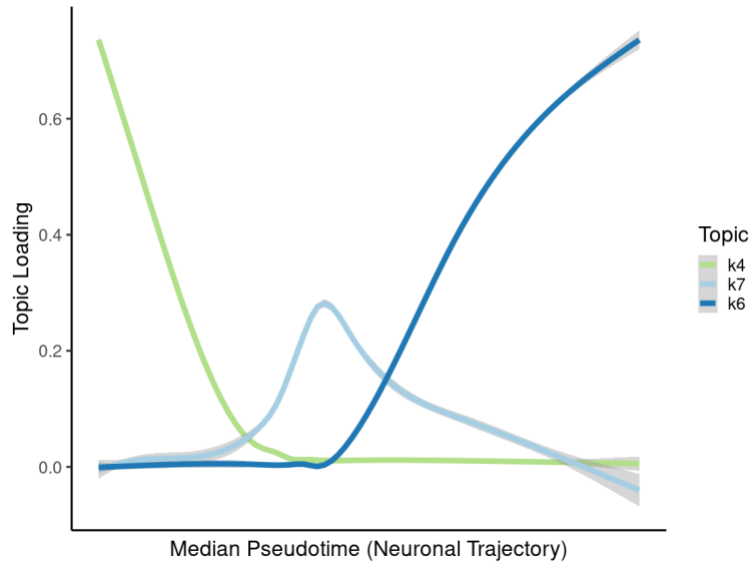


Figure S6. Relationship Between Topics and Neuronal Trajectory Pseudotime, Related to Figure 4. Comparing pseudocell topic loadings to median pseudotime values along the neuronal trajectory highlight topics corresponding to early, intermediate, and late stages of neuronal differentiation.

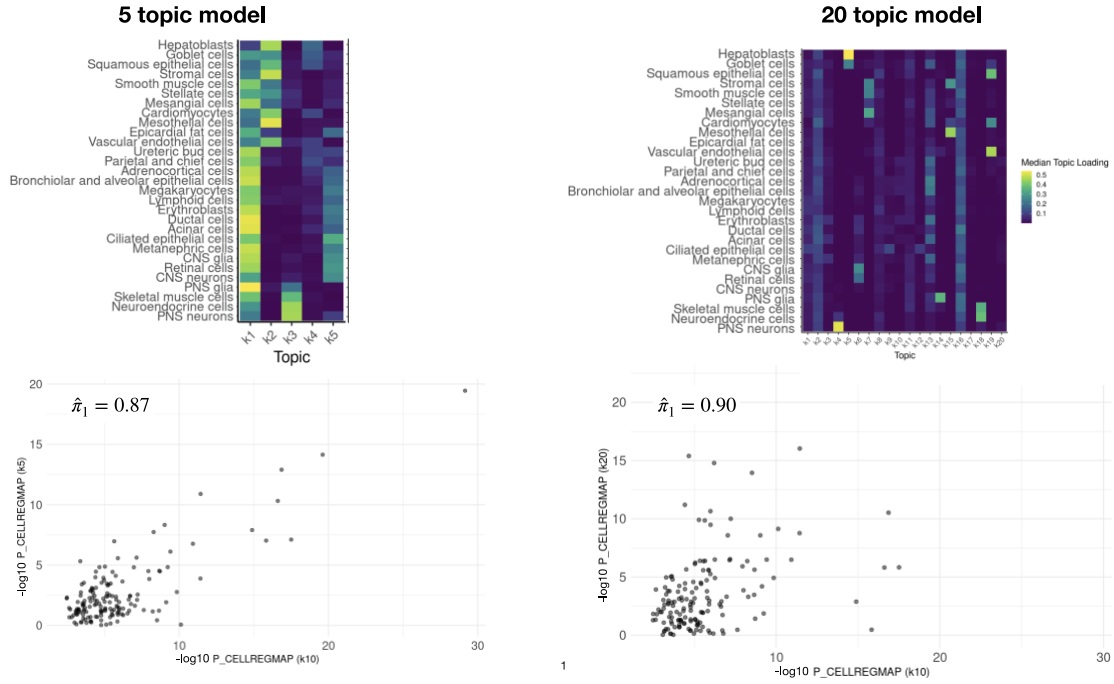


Figure S7. Cell Type Loadings of Topic Models with Different K, Related to STAR Methods. (Top) Heat map of median topic loadings in each cell-type for a 5-topic and 20-topic model. (Bottom) Comparison of $-\log_{10}$ P-values of reported significant topic eQTLs discovered using a 10-topic model (x-axis), to the $-\log_{10}$ P-values obtained when using a 5-topic or 20-topic model. Inset shows estimated replication rate, $\hat{\pi}_1$.

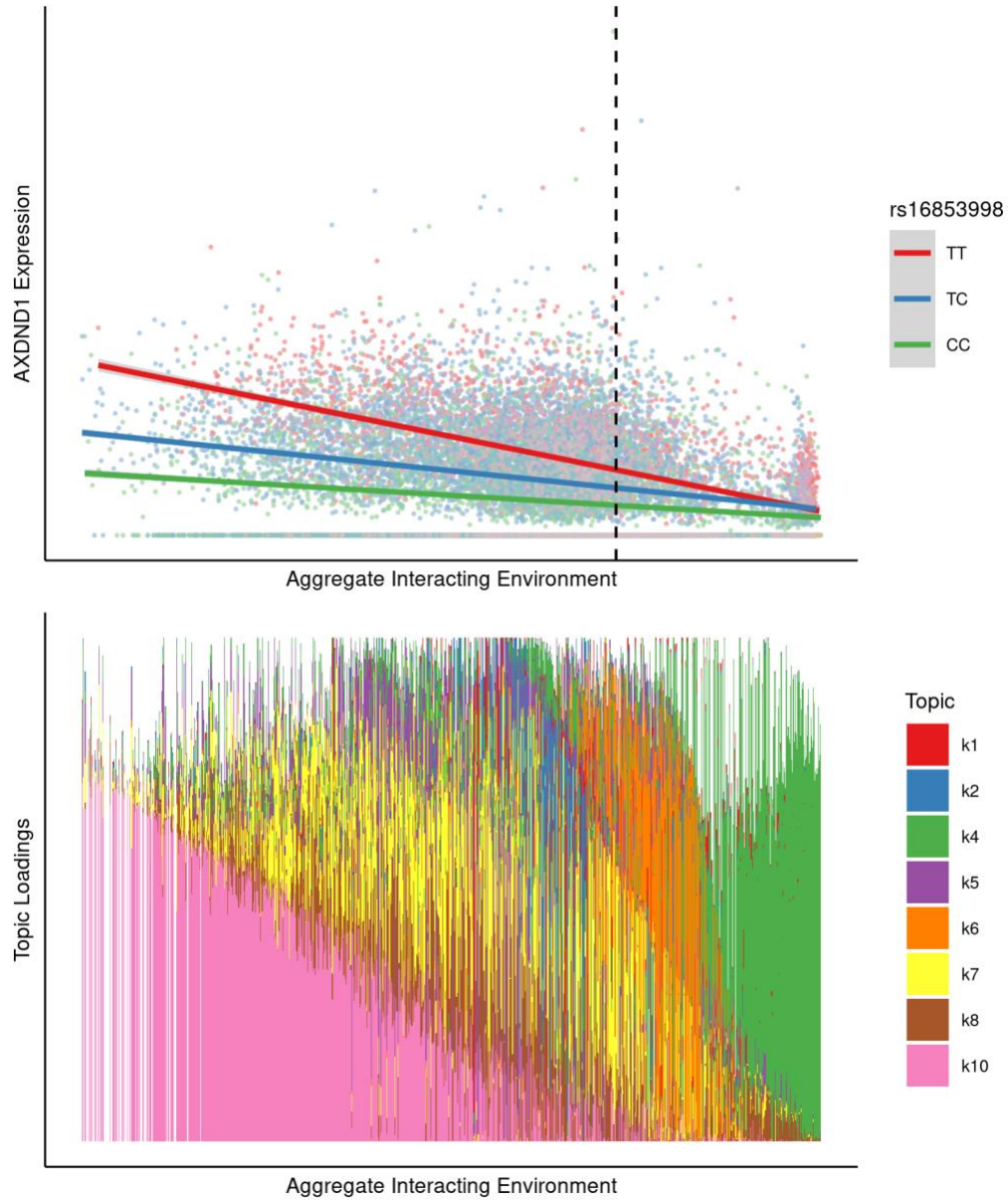


Figure S8. *AXDND1* Topic eQTL, Related to Figure 4. Example topic eQTL for the gene *AXDND1* with maximal effect in cells highly loaded for topic 10 (pink), a topic associated with a ciliary gene program.

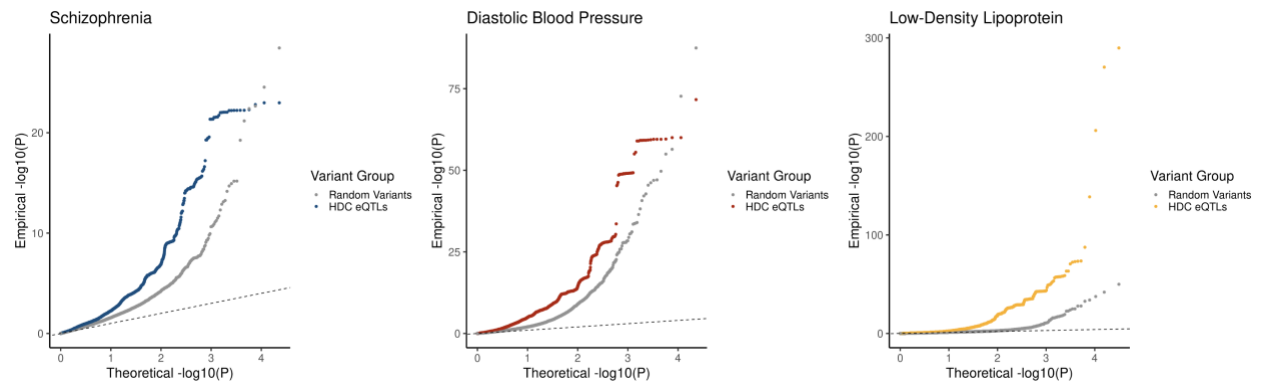


Figure S9. HDC eQTL Enrichment for Trait Association, Related to Figure 5. HDC eQTLs (colored) display inflation of small p-values from the GWAS for schizophrenia (left), diastolic blood pressure (center), and low-density lipoprotein (right), compared to random SNPs (gray).