

Supplementary Methods

Genetic ancestry assessment

To assert ancestry of the LORD cohort, we performed a PCA analysis of the genotypes using plink V1.9 software, together with GTEx participants with Lung expression data and reference genotypes from HapMap.

Deconvolution

We downloaded the Human Lung Cell Atlas database, which contains single cell data for 587,000 cells, including gene expression values and curated cell type annotations [1]. We kept only the top 25% most variable genes. Then, we randomly kept 1000 cells for each unique cell type at each annotation level (levels 1, 2 and 3). We verified that this subsampling strategy produced consistent results among replicate annotation runs. Then, in our own gene expression RNAseq data (the LORD cohort, TPM values of gene expression), we subsampled the same genes (top 25% most variable in the reference single cell data). We also removed any gene, whose mean across the cohort was less than 0.001 TPM.

We then used the MuSiC algorithm [2] to deconvolute samples based on our reference dataset and identify relative cell type proportions per sample. MuSiC (Multi-Subject Single Cell deconvolution) is a method for estimating cell type proportions in bulk RNA-seq data using single-cell RNA-seq as a reference and a weighted regression to estimate proportions.

To verify Δ -eQTLs accounted for changes in cellular populations between lung and tumor, we used the results of this analysis to assess the level of collinearity between Peer Factors and the deconvoluted cell type proportions. To do so, we built a multivariate model using the covariates

from the eQTL model as predictive variables for cell type proportions. We then scrutinized the r^2 values of all these models. For the changes in proportions, we considered the paired difference in proportion as the response variable and used the Peer Factors derived from differences of expression as predictive variables, along with the other covariates.

Supplementary Results

Genetic Ancestry of the LORD cohort

Visualization of the PCA of genotypes from the LORD cohort to other resources showed a consistent overlap between LORD participants and individuals from European ancestry.

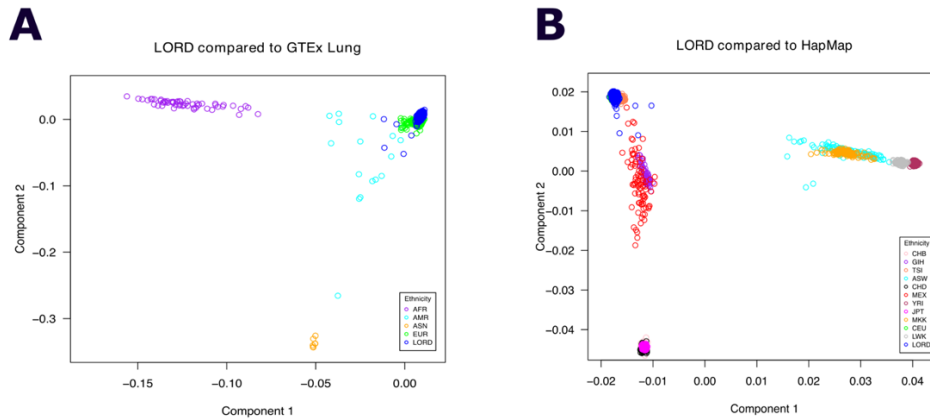


Figure 1: A. PCA plot comparing the genetic ancestry of GTEx to LORD participants. B. PCA plot comparing the ancestry from HapMap to LORD participants.

Evaluation of collinearity between PEER factors and cell type proportion changes

We found out Peer factors account for cell type proportions in each tissue compartment, including the difference of cell type fractions. Indeed, in both lung, tumor and Δ compartments, adjusted r^2

values exceed 0.8 for most cell types in annotation level 1 and 0.6 in annotation level 2 and 3 (Tables 1 to 3).

cell_type	lung	tumor	difference
Endothelial	0.93	0.72	0.85
Epithelial	0.93	0.89	0.88
Immune	0.97	0.85	0.89
Stroma	0.82	0.81	0.79

Table 1: Adjusted r^2 values of the covariate model for annotation level 1.

cell_type	lung	tumor	difference
Airway epithelium	0.94	0.86	0.85
Alveolar epithelium	0.93	0.68	0.80
Blood vessels	0.93	0.75	0.86
Fibroblast lineage	0.91	0.76	0.78
Lymphatic EC	0.29	0.50	0.39
Lymphoid	0.81	0.86	0.82
Myeloid	0.96	0.83	0.90
Smooth muscle	0.81	0.60	0.64
Submucosal gland	0.45	0.25	0.16

Table 2: Adjusted r^2 values of the covariate model for annotation level 2.

cell type	lung	tumor	difference
AT1	0.90	0.75	0.76
AT2	0.90	0.53	0.72
B cell lineage	0.85	0.89	0.89
Fibroblasts	0.88	0.74	0.79
Fibromyocytes	0.70	0.60	0.58
Macrophages	0.93	0.75	0.83
Multiciliated lineage	0.97	0.82	0.91

Table 3: Adjusted r^2 values of the covariate model for annotations from level 3 for which median frequency across the 1030 samples was > 0.01 .

Bibliography

1. Sikkema L, Ramírez-Suástegui C, Strobl DC, Gillett TE, Zappia L, Madisson E, et al. An integrated cell atlas of the lung in health and disease. *Nature Medicine* 2023 29:6 [Internet]. Nature Publishing Group; 2023 [cited 2026 Apr 21];29:1563–77. <https://doi.org/10.1038/s41591-023-02327-2>
2. Fan J, Lyu Y, Zhang Q, Wang X, Li M, Xiao R. MuSiC2: cell-type deconvolution for multi-condition bulk RNA-seq data. *Brief Bioinform* [Internet]. Oxford Academic; 2022 [cited 2026 Apr 19];23. <https://doi.org/10.1093/BIB/BBAC430>