Description of Additional Supplementary Information Files

File Name: Supplementary Data 1

Description: Comparison of number of eGene, vGene, and residGene. The first three columns means whether the gene has significant eQTL, veQTL, or residQTL. The count means how many genes belong to this category. The sum of count for each cell type may not be equal to the number in Table 1 because the residuals of a few genes are NAs so they are not tested in the residual eQTL analysis.

File Name: Supplementary Data 2

Description: Summary of genes with eQTL, cvQTL, and vmrQTL and their overlap. The cv stands for coefficient of variance and vmr stands for variance-mean ratio. The overlap_e_cv indicates the overlap between eGene and cvGene. The overlap_e_vmr indicates the overlap between eGene and vmrGene.

File Name: Supplementary Data 3

Description: Summary of significant deQTL across 14 cell types in OneK1K cohort. Columns J-O represents the TensorQTL results for dispersion eQTL test. pval_nominal means the raw p-value for linear regression. slope means the estimate of beta effect for the linear regression. slope_se means the standard error of the slope estimate. pval_perm indicates the empirical permutation p-value for the top variant of each gene. pval_beta indicates the p-value after beta approximation. qval is the false discovery rate (q-value) computed from pval_beta across all tested genes using the Storey method. The columns P-X represents the slope, slope_se, pval_nominal, for the test of mean, variance, mean-variance residual based on the suffix of _e, _v, and _r. All the tests are two-sided.

File Name: Supplementary Data 4

Description: Overlap between significant residGene and dGene. Jaccard Index is calculated as the number in both sets / the number in either set, i.e., column overlap divided by column union.

File Name: Supplementary Data 5

Description: TensorQTL results using the residuals from the mean regressed out of the dispersion. phenotype_id : the phenotype being tested, i.e., gene name

num_var : the total number of cis-variants tested within the defined window for this phenotype

beta_shape1 : the first shape parameter (α) of the Beta distribution fitted to the permuted minimum nominal p-values for this phenotype; used to compute pval_beta

beta_shape2 : the second shape parameter (β) of the Beta distribution fitted to the permuted minimum nominal p-values for this phenotype; used together with beta_shape1 to parameterise the null distribution

true_df : the effective degrees of freedom for the two-sided t-test after accounting for the variance explained by covariates included in the model; reflects the reduction in sample size due to covariate correction

pval_true_df : the nominal p-value for the top variant recomputed using the true (covariate-corrected) degrees of freedom rather than the naive n minus 2

variant_id : the top associated cis-variant for this phenotype, i.e. the variant with the smallest nominal p-value in the cis window

start_distance : signed distance in base pairs from the variant to the start coordinate of the phenotype (e.g. TSS for genes); negative values indicate the variant is upstream

end_distance : signed distance in base pairs from the variant to the end coordinate of the phenotype (e.g. TES for genes); negative values indicate the variant is upstream of the end

ma_samples : the number of samples carrying at least one copy of the minor allele

ma_count : the total count of minor alleles summed across all samples (i.e. twice the number of homozygous carriers plus the number of heterozygous carriers)

af : allele frequency of the alternative allele in the study sample

pval_nominal : the nominal p-value from the linear regression t-test for the association between genotype and phenotype for the top variant; not corrected for multiple testing across variants or phenotypes

slope : the effect size (β coefficient) from the linear regression, representing the change in phenotype per additional copy of the alternative allele

slope_se : the standard error of the slope estimate

pval_perm : the empirical permutation p-value for the top variant, defined as the proportion of permutations in which the minimum nominal p-value across all cis-variants was equal to or smaller than the observed pval_true_df

pval_beta : a more precise version of pval_perm obtained by evaluating pval_true_df against the fitted Beta(α, β) null distribution rather than relying on the discrete permutation count; this is the per-gene adjusted p-value used for downstream FDR control

qval : the false discovery rate q-value computed from pval_beta across all tested phenotypes using the Storey method; the standard threshold of 0.05 is used to define genome-wide significant QTLs

File Name: Supplementary Data 6

Description: FUMA pathway enrichment results. FUMA's GENE2FUNC gene-set/pathway enrichment uses a hypergeometric (over-representation) test against curated gene sets, which is an one-sided test. N = the number of genes in the tested gene set (within FUMA's background gene list) and n = the number of your input genes that overlap that gene set; the P-value is the upper-tail probability of observing at least the observed overlap between input gene list and a gene set by chance, given a fixed background gene list; The adjusted P-value is computed with the Benjamini–Hochberg (BH) FDR procedure applied per gene-set data source/category (e.g., GO biological processes, KEGG pathways, hallmark sets); The "genes" column lists exactly those overlapping input genes. The GO biological process results were only shown with the top 15 enriched pathways.

File Name: Supplementary Data 7

Description: Trans effect of 65 candidate deQTL-dGene pairs. The columns indicate gene id, top SNP id, A1, A2, allele frequency of A1, cell type, number of trans-eQTL, number of trans-veQTL, and number of trans-deQTL.

File Name: Supplementary Data 8

Description: Interaction test results for candidate deQTLs. Interaction term, the covariate that is used for the interaction test. Cell_type, the cell type in which the interaction analysis was performed. phenotype_id, the phenotype being tested (e.g., gene name); variant_id, the top cis-variant tested; af,

allele frequency of the alternative allele; pval_g / b_g / b_g_se, p-value, effect size (β), and standard error for the main genotype effect (β_g): the average association between genotype and phenotype, independent of the interaction term; pval_i / b_i / b_i_se, p-value, effect size, and standard error for the main effect of the interaction variable (β_i): how much E alone influences the phenotype, regardless of genotype; pval_gi / b_gi / b_gi_se, p-value, effect size, and standard error for the GxE interaction term (β_gi): the primary term of interest, whether the genotype effect on phenotype differs as a function of E; tests_emt, the effective number of independent tests (M_eff) estimated for this gene's cis- window using an eigenMT-style approach on the LD matrix of tested variants; pval_emt, the interaction p-value corrected for multiple variant testing within the cis window, approximately min(1, best pval_gi × tests_emt); analogous to the permutation-based pval_beta in standard cis-QTL output but computed analytically; pval_adj_bh, Benjamini–Hochberg FDR applied across all phenotypes using pval_emt, the genome-wide significance threshold for calling significant interaction QTLs. All the tests are two-sided

File Name: Supplementary Data 9

Description: Threshold of intra-mean for dispersion estimate inflation. The columns indicate the cell type, number of donors tested in the cell type, median of number of cells per donor, and intra-individual mean expression level to remove genes with inflated dispersion estimates.

File Name: Supplementary Data 10

Description: The special case of estimating dispersion using moment method when mean is small.

Pattern # : index of special case based on the simulation of 10,000 individuals, each with 1,000 cells, and all the within-individual distributions are drawn from $NB{\sim}(\mu = 0.001, \theta = 1)$

Nr of 0 (m): number of cells with 0 count of the gene

Nr of 1 (k): number of cells with 1 count of the gene

Nr of 2 (j): number of cells with 2 count of the gene

Nr of individual (N): number of individuals fits this pattern out of 10,000 individuals

Mean: formula of calculating the mean using m, k, and j

Variance: formula of calculating the variance using m, k, and j

Mean estimate: actual estimate of the mean

Variance estimate : actual estimate of the variance

Theta moment estimate: moment estimate of the theta, i.e., dispersion parameter.