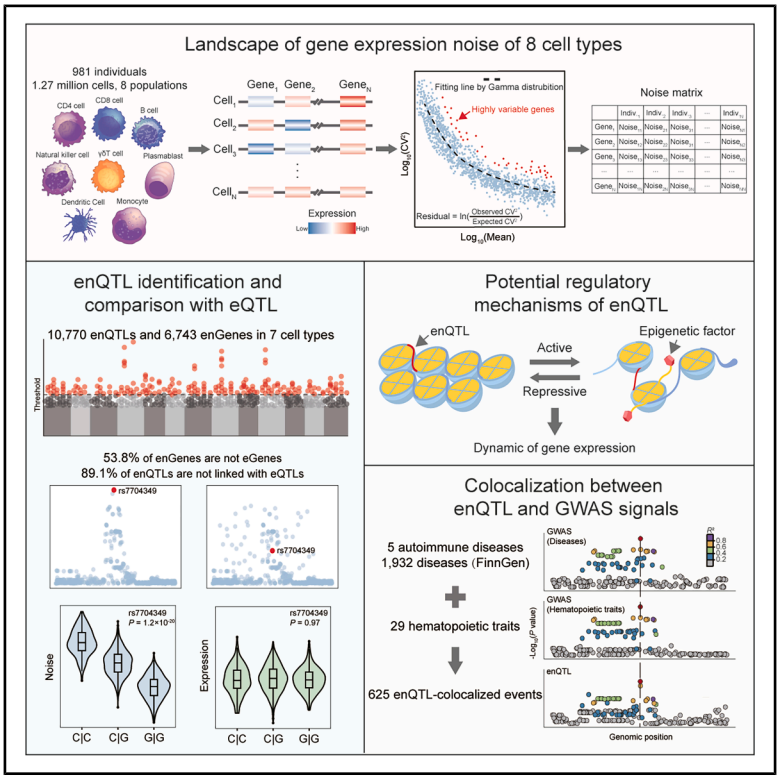
Cell Reports

Article

Genetic determinants of gene expression noise and its role in complex trait variation

Graphical abstract



Highlights

- Atlas of transcriptional noise across eight immune cell types reveals age/sex-specific patterns
- A total of 10,770 enQTLs identified, regulating noise for 6,743 enGenes across 7 cell types
- enQTLs distinct from eQTLs, enriched in repressive chromatin and intergenic regions
- enQTL-GWAS colocalization prioritizes causal variants for hematopoietic traits and diseases

Authors

Yuexuan Long, Xiaolin Ni, Tingwei Chen, ..., Jiyeon Choi, Tongwu Zhang, Erping Long

Correspondence

xiaolin_ni7@126.com (X.N.), tongwu.zhang@nih.gov (T.Z.), erping.long@ibms.pumc.edu.cn (E.L.)

In brief

Long et al. mapped expression noise QTLs (enQTLs) across immune cell types in an atlas of 1.23 million cells. These enQTLs are largely independent of eQTLs, exhibit cell-type-specific regulatory mechanisms, and colocalize with GWAS loci for hematopoietic traits and diseases, implicating noise as an underappreciated molecular mediator of complex traits.

Long et al., 2025, Cell Reports *44*, 116612
December 23, 2025
https://doi.org/10.1016/j.celrep.2025.116612

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Genetic determinants of gene expression noise and its role in complex trait variation

Yuexuan Long,[1,6] Xiaolin Ni,[1,6,7,*] Tingwei Chen,[1,6] Xianzhe Huang,[1] Qiyang Hong,[1] Jixin Wang,[1,2] Cong Wang,[1] Zigeng Huang,[1] Haiqing Xu,[3] Mengyi Sun,[4] Junling Pang,[1] Jiyeon Choi,[5] Tongwu Zhang,[5,7,*] and Erping Long[1,7,8,*]

[1]State Key Laboratory of Respiratory Health and Multimorbidity, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China
[2]School of Medicine, Tsinghua University, Beijing, China
[3]Department of Biology, Stanford University, Stanford, CA, USA
[4]Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA
[5]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA
[6]These authors contributed equally
[7]Senior author
[8]Lead contact
*Correspondence: xiaolin_ni7@126.com (X.N.), tongwu.zhang@nih.gov (T.Z.), erping.long@ibms.pumc.edu.cn (E.L.)
https://doi.org/10.1016/j.celrep.2025.116612

SUMMARY

Even genetically identical cells in homogeneous environments exhibit heterogeneous mRNA abundance, typically called "gene expression noise," which is involved in key cellular activities, evolutionary processes, and disease mechanisms. However, determinants of the gene expression noise and its functional role in variations of human complex traits remain largely unexplored. Here, we established an atlas of gene expression noise from 1.23 million human peripheral blood cells of 981 individuals, identifying its age- and gender-dependent pattern. We then identified 10,770 independent expression noise quantitative trait loci (enQTLs) for 6,743 unique enGenes across seven immune cell types. Most enQTLs were distinct from expression quantitative trait loci (eQTLs) and showed differential enrichment of functional elements across the genome. Co-localization of enQTLs with trait-associated genetic loci interpreted previously unexplained loci. Overall, this study unravels the genetic determinants of gene expression noise and implicates as a previously underappreciated mechanism underlying variation of human complex traits and diseases.

INTRODUCTION

The intrinsic stochasticity of biochemical reactions contributes to a wide distribution of gene expression across a seemingly homogeneous population of cells. This phenomenon of transcriptional variability, called "gene expression noise," has been widely observed in prokaryotic and eukaryotic systems.[1] Classically, noise can be quantified by time-resolved reporter assay of individual genes.[2] More recently, single-cell techniques, along with statistical approaches, have been applied to distinguish true biological cell-to-cell variability from measurement errors.[3] Fueled with these methodological advances, several genetic and epigenetic features were identified to be the potentially deterministic components of noise. In budding yeast, less-active genes with high nucleosome occupancy close to the transcriptional start site (TSS) and genes with presence of TATA box in their promoter sequences are associated with elevated noise.[4,5] In mammalian cells, the association between TATA box and noise has been consistently reported.[6] Moreover, genes with the absence and shorter length of promoter-region CpG islands were associated with attenuated noise.[7,8] However, our understanding of the interindividual determinants (e.g., the genetic variants across human genome) of noise remains unknown.

Gene expression noise is suggested to be involved in a variety of important biological and pathological processes. For example, noise is elevated in early pluripotent cells and attenuated in later embryonic development stages.[9–11] In terms of senescence, transcriptional variability increases with age, and the disruption of noise control may lead to abnormal immune responses.[12] Moreover, cell cycle and gene expression noise were intimately connected, as the genes regulating cell cycle were variably expressed across different phases but stable within the same phase.[13] Most human phenotypes and diseases are complex traits that are contributed by genetic variants from numerous loci across genome[14] with mediators of molecular traits.[15] However, current studies of molecular quantitative trait loci (QTLs) mainly focused on steady-state average abundances of a trait (e.g., gene expression); whether and to what extent the variability of a trait (e.g., noise) may contribute to human complex traits remains to be explored.

To address the above fundamental questions, we leveraged the population-scale single-cell datasets of human peripheral blood cells to characterize the transcriptome-wide noise atlas.

We then identified 10,770 independent expression noise QTLs (enQTLs) across seven immune cell types, which were largely independent from expression QTLs (eQTLs) and can explain a substantial proportion of previously unexplained immune trait-associated loci from genome-wide association studies (GWASs), as shown by 625 enQTL-colocalized events. Overall, this study demonstrated the genetic architecture of gene expression noise and highlighted the important role of noise underlying genetic associations with complex traits.

RESULTS

Gene expression noise across cell types and associations with age and gender

We curated a single-cell transcriptome atlas from the OneK1K cohort, comprising 1,233,834 peripheral blood mononuclear cells (PBMCs) from 981 healthy individuals of northern European ancestry.[16] To accurately calculate the gene expression noise by using a sufficiently high cell number per individual, we grouped the cells into eight major cell types based on established annotations: $CD4^+$ T (CD4) cells, $CD8^+$ T (CD8) cells, natural killer (NK) cells, B cells, monocytes (Mono), gamma-delta T (gdT) cells, dendritic cells (DCs), and plasmablasts (plasma) (Figure S1A; Table S1).[16] The numbers of cells in each cell type ranged from 3,754 (0.3% of the total cells) for plasma cells to 624,570 (50.6%) for CD4 cells (Figure S1A). The dimensionality reduction of uniform manifold approximation and projection (UMAP) and clustering analysis revealed a clear hierarchical architecture among T (CD4, CD8, and gdT), NK, B, DC, Mono, and plasma cells (Figure S1B). Additionally, we observed cell-type-specific expression of canonical markers for all cell types, reassuring the validity of cell-type annotation (Figure S1C). On average, 1,255 cells were included per individual and 71.7% (703/981) of individuals contain all eight cell types across the cohort (Figures S1D and S1E; Table S2).

Technical noise is dependent on the average read count of a gene (i.e., mean-variability dependence) in single-cell data.[17] To deconvolute the biological cell-to-cell noise from technical noise, we applied a statistical approach addressing mean-variability dependence of over-dispersed distribution. Namely, we fit the mean and squared coefficient of variation (CV^2 excluding low-expressed genes with extremely high CV^2) based on gamma distribution for each individual and calculated the normalized gene expression noise based on residual and logarithmic transformation for each cell type (Figure 1A, STAR Methods).[3] We observed the significantly positive correlation between cell numbers and fitting performance (correlation coefficient = 0.84, $p \leq 2.2 \times 10^{-26}$) and retained the individuals with robust fitting performance (adjusted $R^2 > 0.6$; Figure S2A. The plasma cells were thus excluded from subsequent analyses because of the limited sample size for further QTL analysis (only 55 individuals with adjusted $R^2 < 0.6$; the rest of the cell types have 242–981 qualifying individuals).[18] The calculated noise exhibited significantly reduced correlation to mean expression in each cell type compared to CV^2, validating the efficiency of deconvoluting the mean-variability dependence (Figure 1B).

We then identified the highly variable genes (HVGs) as the genes exhibiting significantly elevated expression noise (false discovery rate [FDR] $\leq$0.05) in at least 10% of individuals in each cell type (Figure S2B; STAR Methods). A total of 437 HVGs were identified across all cell types, ranging from six HVGs in DC cells to 350 HVGs in CD4 cells (Figures 1C and S2C; Table S3). A substantial proportion of HVGs (41.2%, 180/437) are shared in at least two cell types, while the remaining (59.8%, 257/437) are cell-type specific (Figure S2C; Table S3). Notably, we identified two genes that were HVGs across all cell types: *GNLY* (relevant to innate immune system)[19] and *CRIP1* (relevant to cell proliferation and growth)[20,21] (Figure S2C). As expected, lineage-specific genes are significantly enriched (all $p \leq 0.01$) in HVGs across cell types, including *GZMK* and *IL7R* in CD4 cells and *XCL1* in NK cells (Figure 1D; Table S4). Moreover, we observed the significant enrichment of HVGs in adaptive-immune-response pathways for CD4, CD8, NK, B, and Mono cells, and cell-cycle pathways for CD4 and NK cells (Figure S2D). Together, these findings, consistent with previous studies,[13,22] implied the involvement of gene expression noise in immune and cell-cycle process.

To assess the potential impact of age, gender, and cell type on gene expression noise, we aggregated the noise of all genes into individual levels based on normalized gene expression noise for each cell type. Our first observation was that the individual-level noise elevated as age increased, and this trend was consistent across six cell types, with the exception of DC cells (Figures 1E and 1F), supplementing the existing evidence of noise-age relationship.[12,23–25] Our second observation is the significantly elevated noise among males relative to females across six cell types (Figure 1G). This trend has been consistently identified in subsets of both autosomal and sex chromosome genes (Figures S3A and S3B). Moreover, we found no significant differences of age between males and females (Figure S3C), and the noise difference between male and female significant persisted after adjusting the age as a covariate (ANOVA, $p = 2.36 \times 10^{-11}$), which confirms that the elevated noise among males is genuine. Interestingly, we noted that monocytes exhibited the highest individual-level noise among cell types (Figure 1G), and monocytes are known for their heterogeneity and plasticity in response to environmental dynamics.[26]

Identification of enQTLs

To investigate how genetic variation influences gene expression noise in a cell-type-specific manner, we developed a computational pipeline to identify the enQTLs and their target genes (enGenes) with genetically driven gene expression noise (STAR Methods). We identified enQTLs using tensor-QTL based on Pearson product-moment correlation and an adaptive permutation scheme (STAR Methods).[27] Within each cell type, we incorporated the following covariates: sex, age, six genotype-based principal components, and a Bayesian factor (probabilistic estimation of expression residuals, separately for each cell type; Figure S4A)[28] to eliminate unwanted global variation that could obscure the subtle effects of *cis*-acting genetic SNPs.

To determine the significance threshold of enQTL, we examined the distribution of effect sizes and nominal p of lead SNPs after permutation, under the assumption that random SNP selection results 50% (null hypothesis) of consistent allelic direction across cell types.[29,30] We selected a p threshold of 0.001, where

CellPress
OPEN ACCESS

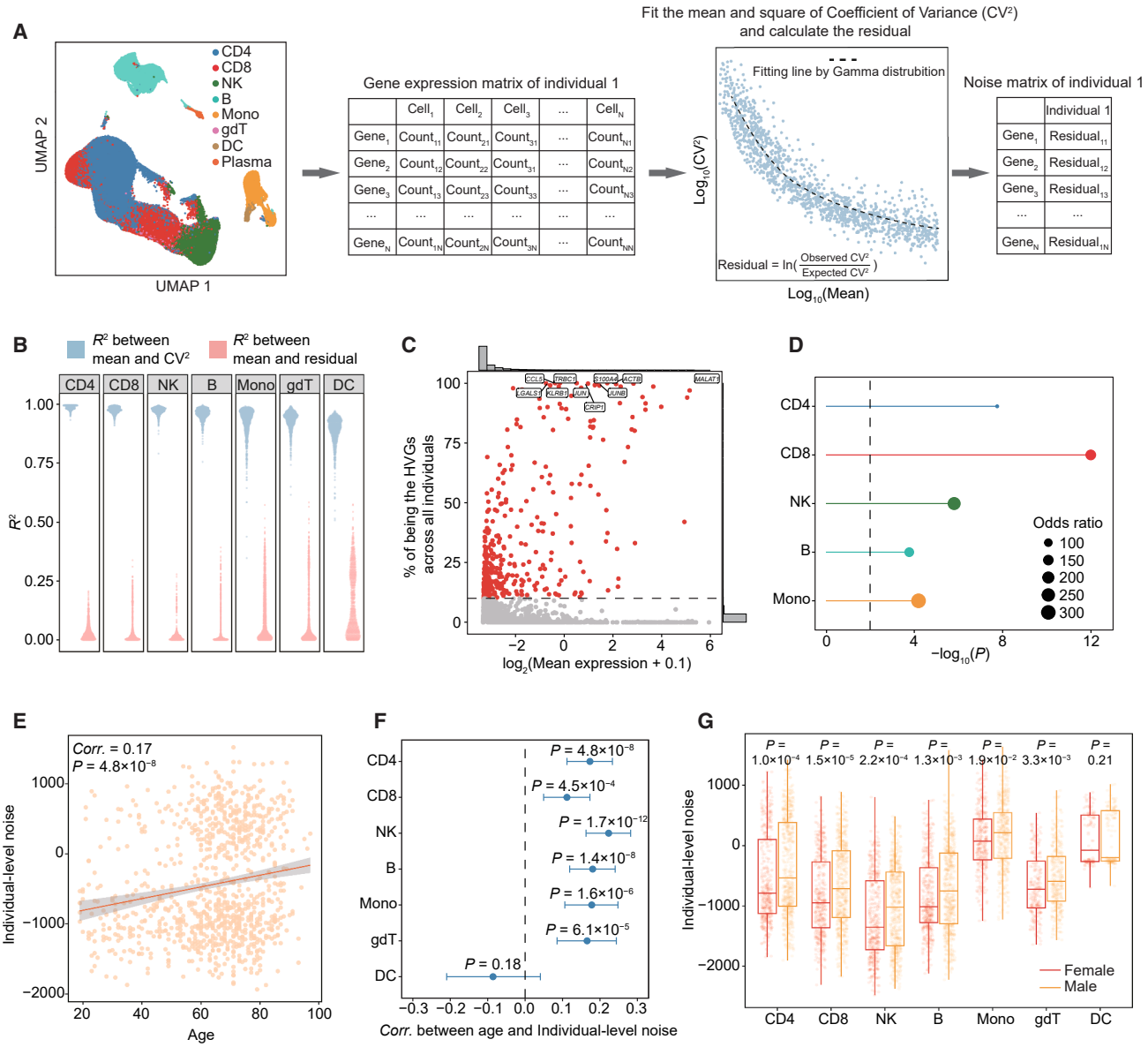


Figure 1. Characterization of gene expression noise, HVGs, and individual-level noise

(A) Schematic overview of the framework used to calculate gene expression noise across cell types. The matrixes left and right represent the gene expression and gene expression noise in one cell type of an individual, respectively. The dot plot on the right is the fitted graph of CV^2 and the mean in one cell type of an individual. A blue dot represents a gene, and the dotted black line represents the gamma distribution fitting.

(B) The square of correlation coefficients between mean expression levels and CV^2 (blue) or residual (red) were calculated using Pearson's product-moment correlation. Each dot represents an individual.

(C) Distribution of rates of being the HVGs across all individuals for all genes in CD4 cells. The gray line marks the 10% threshold for identification of HVGs. Red dots represent HVGs, while gray dots indicate non-HVGs. The labeled genes are the top 10 genes with the highest rates of being HVGs across all individuals. The bar graph shows the distribution of mean expression (*x* axis) and highly expression noise rates (*y* axis).

(D) The enrichment of lineage-specific genes in HVGs across five cell types with known lineage-specific gene information. The enrichment *p* value was determined by Fisher's exact test.

(E) Correlation between age and the sum of residuals in CD4 cells, with *p* value and correlation coefficient calculated by the Pearson's product-moment correlation. Each dot represents an individual. The gray shade indicates the 95% confidence interval.

(F) Pearson correlation coefficients with 95% confidence intervals for the relationship between age and sum of residuals across seven cell types: CD4 (*N* = 981), CD8 (*N* = 981), NK (*N* = 978), B (*N* = 973), Mono (*N* = 716), gdT (*N* = 580), and DC (*N* = 242) cells.

(G) Comparison of the individual-level noise between males and females across seven cell types: CD4 (*N* male = 565, *N* female = 416), CD8 (*N* male = 565, *N* female = 416), NK (*N* male = 562, *N* female = 416), B (*N* male = 565, *N* female = 408), Mono (*N* male = 397, *N* female = 316), gdT (*N* male = 342, *N* female = 238), and DC (*N* male = 129, *N* female = 113) cells.

p values were calculated by the two-side Wilcoxon signed-rank test.

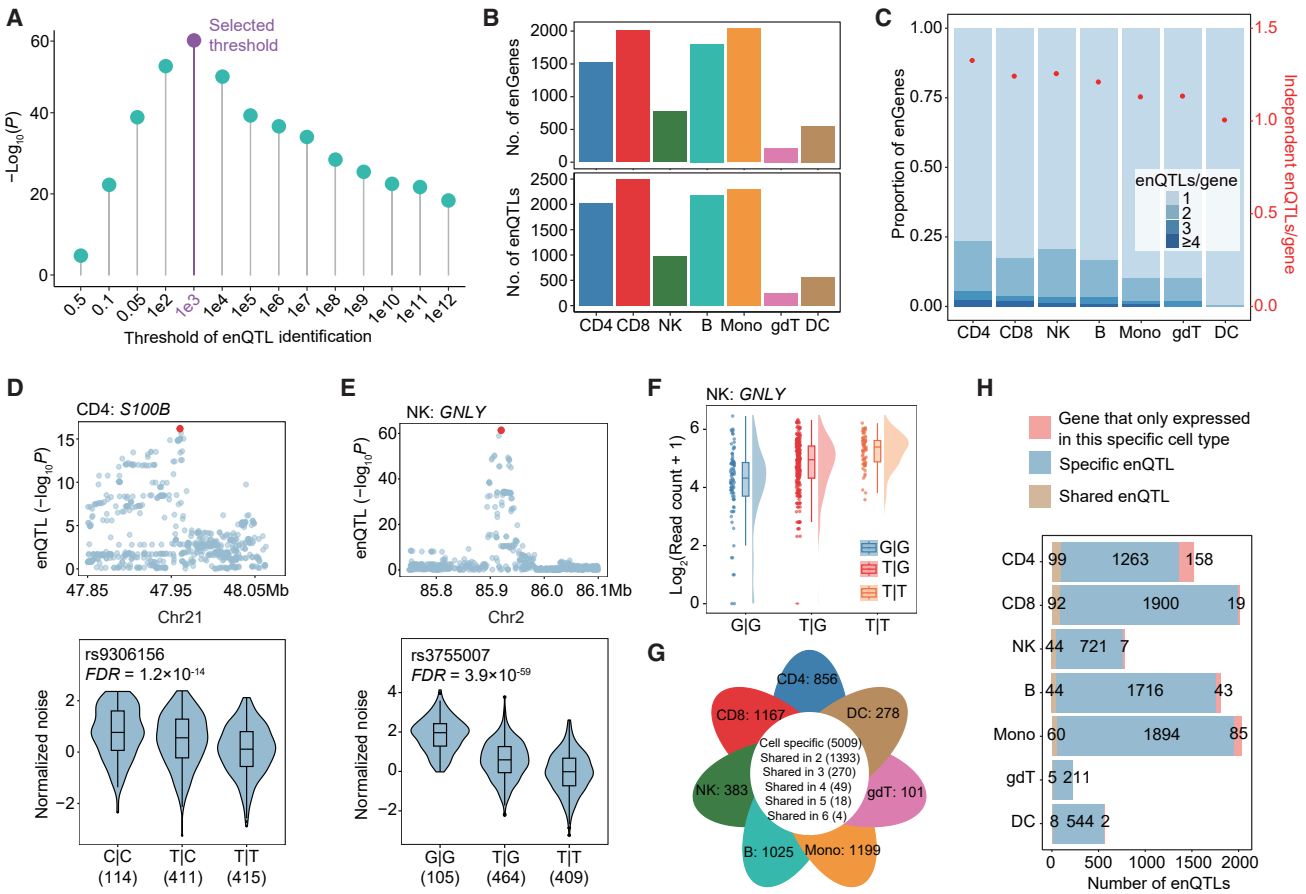


Figure 2. Identification of enQTL across cell types

(A) Statistical difference between enQTL directional consistency and null hypothesis (50%) between any of two cell types across a series of thresholds. The threshold of 0.001 was selected (purple). Binomial test was used to determine statistical difference.

(B) The number of enGenes (top) and enQTLs (bottom) identified across seven cell types.

(C) Allelic heterogeneity of enQTLs is depicted by two measures: the proportion of enGenes with one or more independent enQTLs (blue stacked bars; left *y* axis) and the mean number of enQTLs per gene (red dots; right *y* axis).

(D and E) Examples of enGenes, including a cell-cycle-related gene *S100B* in CD4 cells (D) and an immune-related gene *GNLY* in NK cells (E). In the Manhattan plot, the red dots represent the lead SNPs, with the sample size of each genotype indicated in brackets. In each box, the central line denotes the median, notches represent the 95% confidence interval, the box indicates the interquartile range and whiskers extend up to 1.5 times the interquartile range.

(F) Distribution of *GNLY* expression (read count) between three rs3755007 haplotypes. Violin and dot plots depict the distribution of $\log_2$(read count + 1) across haplotypes of single cells. In each box, the central line denotes the median, notches represent the 95% confidence interval, the box indicates the interquartile range and whiskers extend up to 1.5 times the interquartile range.

(G) The number of shared and cell-type-specific enGenes identified across the seven cell types. The numbers on the petals represent the number of cell-type-specific enGenes for each cell type.

(H) The number of shared and cell-type-specific enQTLs across the seven cell types. The red segment represents cell-type-specific enQTLs caused by genes only expressed in one cell type.

the allelic direction consistency of SNPs most significantly deviated from null hypothesis ($p = 3.5 \times 10^{-58}$, binomial test) with 89.7% of the SNPs showing a consistent direction (Figure 2A). A total of 10,770 independent enQTLs involving 6,743 unique enGenes across seven cell types (Figure 2B) were identified, with breakdowns as follows: 2,016 enQTLs and 1,520 enGenes in CD4 cells, 2,497 enQTLs and 2,011 enGenes in CD8 cells, 969 enQTLs and 772 enGenes in NK cells, 2,182 enQTLs and 1,803 enGenes in B cells, 2,304 enQTLs and 2,039 enGenes in Mono cells, 245 enQTLs and 216 enGenes in gdT cells, and 557 enQTLs and 554 enGenes in DCs. Notably, 120 enGenes were identified as HVGs, including immune-function genes (e.g., *KRT1* and *GNLY*)[31,32] and cell-cycle genes (e.g., *S100B*)[33] (Figures 2D and 2E; Tables S5 and S6). Expression distribution of *GNLY* showed that the noise-increasing allele (G) is associated with both lower expression and greater variability (Figure 2F), consistent with the haplotype analysis (Figure 2E). We performed conditional enQTL identification for five rounds, where lead SNP-gene pairs from earlier rounds were included as covariates in subsequent rounds (Table S7; STAR Methods). The majority (5,983/6,743, 88.7%) of enGenes had a single independent enQTL, while 0.5% (3/554 in DC) to 23.6%

(359/1,520 in CD4) of enGenes had multiple independent *cis*-enQTLs, indicating the allelic heterogeneities in these enGenes (Figures 2C and S4B). To account for the effect of cellular heterogeneity on enQTLs, we performed enQTL analysis separately across 14 distinct, highly homogeneous subtypes (cell number >10,000), identifying 233–1,004 enGenes and 285–1,133 enQTLs (Tables S8 and S9).

To validate the reproducibility of enQTLs, we focused on an independent cohort of 149 Japanese individuals from the Asian Immune Diversity Atlas (AIDA)[34] and tested whether the allelic effect direction was concordant between cohorts (Figure S4C; STAR Methods). A total of 16 testable enQTLs were included if they were (1) significant (FDR < 0.001) in the original OneK1K dataset (n = 981), (2) located in the exonic or UTR regions, and (3) heterozygous individuals and their enGenes had non-all-zero expression in AIDA cohort. Our results showed that all enQTLs exhibited concordant direction of the noise effect between the AIDA and OneK1K cohorts. Notably, despite reduced sample size in the AIDA cohort, four of 16 enQTLs remained statistically significant in allele-specific noise effects ($p < 0.01$), as exemplified by enQTL lead SNP rs1131017 and enGenes *RPS26* in five OneK1K-significant cell types (Figure S4D).

Of all 6,743 enGenes, 5,009 genes were cell-type-specific enGenes (Figure 2G). Notably, we identified four enGenes shared in six cell types (*HLA-C*, *HLA-DQB1*, *EIF5A*, and *RPS26*[35–38]), which were reported to be involved in immunity and cell cycling (Figure S5A). The lead SNPs targeting the same enGenes in different cell types could be genetically linked. Therefore, we applied multivariate adaptive shrinkage (MASH) to systematically test the shareability of enQTL across cell types.[39] We found that 315 enQTLs (3.5%) were shared in at least two cell types (Table S10). In addition, only a small fraction of the cell-type-specific enQTLs (314, 3.5%) were associated with the enGenes that were specifically expressed in a single cell type (Figure 2H).To investigate mechanisms underlying cell specificity, we investigated transcription factor (TF)-mediated regulation via motif disruption analyses and hierarchical linear modeling. This revealed 1,083 enQTL-TF-enGene regulatory trios, including rs2855990-*STAT4*-*TRBC1* and rs2566142-*SMARCC1*-*TMEM156* (Figures 5B and 5C), where TF expression correlates with enGene noise only in enQTL-significant cell types. Together, these results indicated a considerable cell-type specificity in enGenes and enQTLs, which could be attributed to context-dependent TF binding.

Independence of enQTLs from eQTLs

To evaluate the relationship between enQTLs and eQTLs, we identified the eQTL using the same post-QC (quality control) data and analytic procedure. We observed that 46.2% (3,122/6,743) of enGenes are eGenes, ranging from 12.3% (68/544 in gdT cells) to 62.4% (949/1,520 in Mono cells) (Figure 3A). Notably, the proportion of overlapping genes showed no significant correlation with the sample sizes in each cell type ($p = 0.23$), suggesting that the overlap is not driven by discovery power. In genes with both an enQTL and an eQTL, most of the lead SNPs (91.8%) for enQTLs and eQTLs were more than 10 kb away from each other (average distance of 462 kb; Figure 3B), and 89.1% of them were not genetically linked (linkage disequilibrium[LD] $R^2 < 0.2$ in 1,000 Genomes Project European [1000G EUR] population; Figure 3C). Among the co-lead SNPs (the same lead SNP targeting the same gene), 63.2% (48/76) had consistent allelic directions (e.g., higher average expression with higher expression noise), while 36.8% (28/76) were discordant (Figures 3D and 3E). We highlighted four different scenarios with specific enQTLs (Figure 3F), specific eQTLs (Figure 3G), co-lead SNPs with consistent direction (Figure 3H), and discordant direction (Figure 3I). Together, these results highlighted the independence of enQTLs from eQTLs, implying that *cis*-regulation of gene expression noise is independent from that of mean expression level.

Functional enrichment and potential mechanisms of enQTLs

To investigate the potential mechanisms of enQTLs in noise regulation, we utilized an annotation tool (ANNOVAR), epigenomic and three-dimensional genomic data from the Encyclopedia of DNA Elements (ENCODE) project, and enhancer data from Super-Enhancer database (SEdb) (Table S11).[40–42] We found that enQTLs are notably enriched in 5′-UTRs for all seven cell types and in exons and upstream/downstream regions of genes for CD4, CD8, and NK (Figure 4A; odds ratio [OR] >1, $p \leq 0.01$). Additionally, enQTLs were enriched in accessible chromatin regions, active histone modifications (such as H3K27ac, H3K4me3, and H3K4me1), enhancers, and regions associated with chromatin structure, including loop anchors and topologically associating domain (TAD) boundaries for all seven cell types (Figure 4B; OR > 1, $p \leq 0.01$). In addition, enQTLs were significantly enriched in regions marked by bivalent modifications in CD4, CD8, B, and Mono (Figure 4B, OR > 1, $p \leq 0.05$). These results support the involvement of enQTL in transcriptional regulation.

Having established that enQTLs are largely distinct from eQTLs, we next explored whether enQTLs and eQTLs display different patterns of functional enrichment. The annotation results from genomic regions revealed that, compared to eQTLs, enQTLs are not significantly depleted in intergenic regions in CD4, CD8, NK, B, and Mono (Figure 4C). In terms of epigenetic modifications, eQTLs are significantly depleted in repressive histone modifications of H3K27me3 and H3K9me3 (OR $\leq$ 1, $p \leq$ 0.05). However, such depletions were not observed for enQTLs, with all ORs $\geq$1 in repressive histone marks (Figure 4C). In general, the enrichment of active regions in enQTLs is attenuated relative to eQTLs (Figure 4C), indicating that repressive chromatin regions and intergenic areas do not adversely affect enQTLs. Together, our data implicated that enQTLs and eQTLs may be mediated by different mechanisms, such as transcription initiation and chromatin remodeling.

enQTLs help interpret GWAS loci

We next asked whether the enQTLs can help uncover the potential mechanisms for trait-associated GWAS loci. We performed the colocalization analysis across 29 hematopoietic traits and five autoimmune diseases, assessing whether the enGene-trait association is driven by the same set of casual SNPs (STAR Methods). In total, we identified 214 enQTL-colocalized events (unique enGene-trait-cell-type pairs) for blood traits and/or

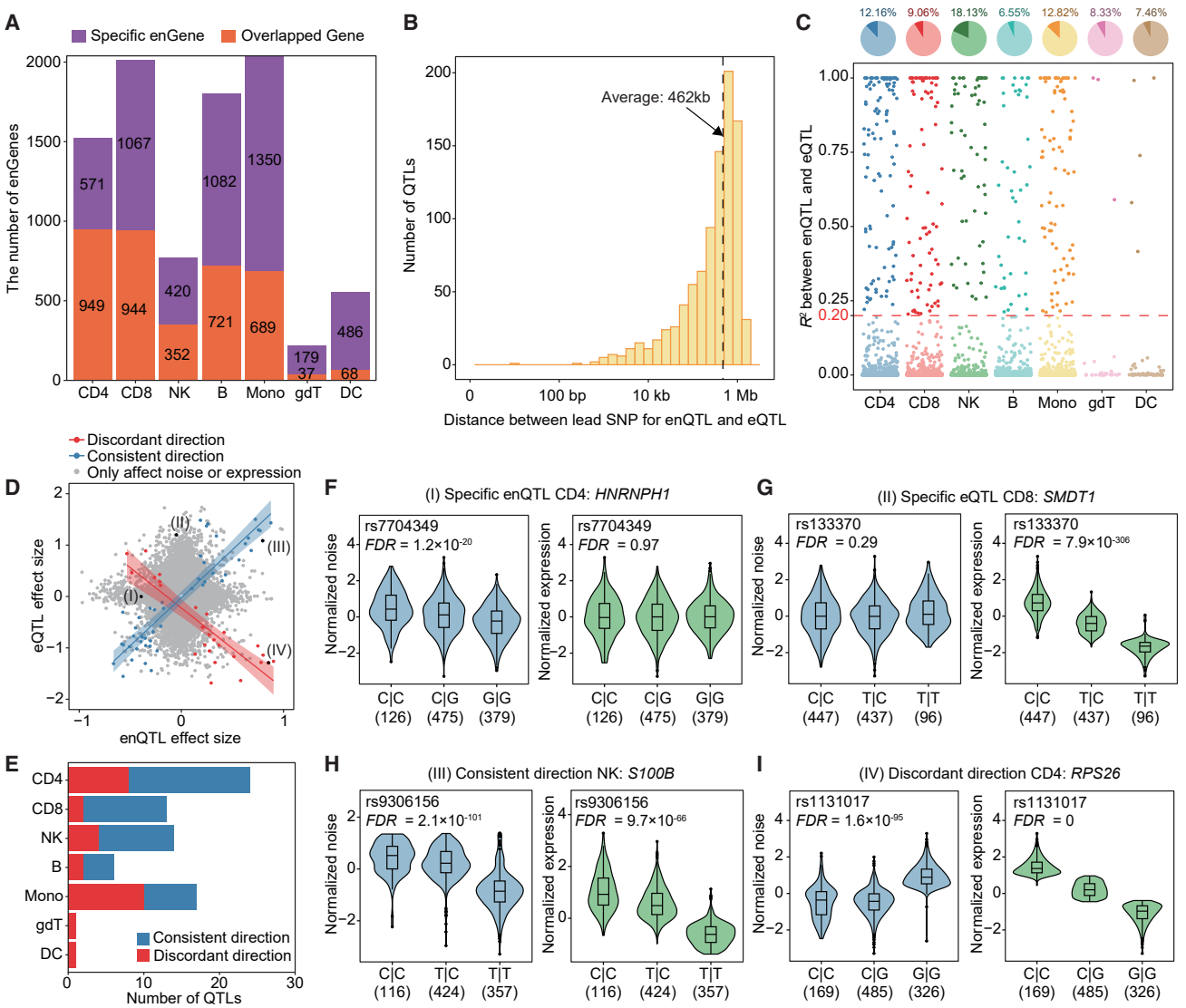


Figure 3. enQTL versus eQTL comparison

(A) The number of specific enGenes (purple) and overlapping genes (orange) across the dataset, showing the distribution of genes regulated by enQTLs, eQTLs, or both.

(B) Distance between lead SNP for enQTL and eQTL of the overlapping genes. The dashed line represents the average distance.

(C) Linkage disequilibrium (LD) R^2 (1000 Genome, phase 3, EUR population) between the lead enQTL and eQTL SNPs for overlapping genes. The pie chart above indicates the proportion of SNP pairs with R^2 values >0.2 and ≤0.2. Dark and light colors represent SNP pairs >0.2 and ≤0.2, respectively.

(D) Scatterplot showing the effect sizes of SNPs on gene expression noise (*x* axis) and gene expression levels (*y* axis), including SNPs that affect noise or expression (gray), or both (blue and red colored).

(E) Number of QTLs affecting both noise and expression with consistent (blue) or discordant (red) directions across cell types.

(F–I) Four examples are highlighted, namely *HNRNPH1* (specific enQTL affecting only gene expression noise), *SMDT1* (specific eQTL affecting only gene expression levels), *S100B* (QTLs with consistent effect direction on noise and expression), and *RPS26* (QTLs with discordant effect directions on noise and expression). The sample size of each genotype is indicated in brackets. In each box, the central line denotes the median, notches represent the 95% confidence interval, the box indicates the interquartile range and whiskers extend up to 1.5 times the interquartile range.

complex diseases (posterior probability of H4 > 0.7; Tables S13, S14, and S15; Figure S6A). The same pipeline was applied to identify 1,473 eQTL-colocalized events (Table S16). We observed that 93.93% (201/214) of the enQTL-colocalized events are independent from eQTL-colocalized events (Figure 5A). Notably, we observed that, in 10 of 12 autoimmune-disease colocalized events, the risk allele was associated with attenuated noise of the target genes (Figure S6B).

There are two cases of missense SNPs that were nominated from enQTL-colocalized events. The first case is rs1801133, a missense variant of *MTHFR* with known associations of reduced enzymatic activity, higher homocysteine levels,

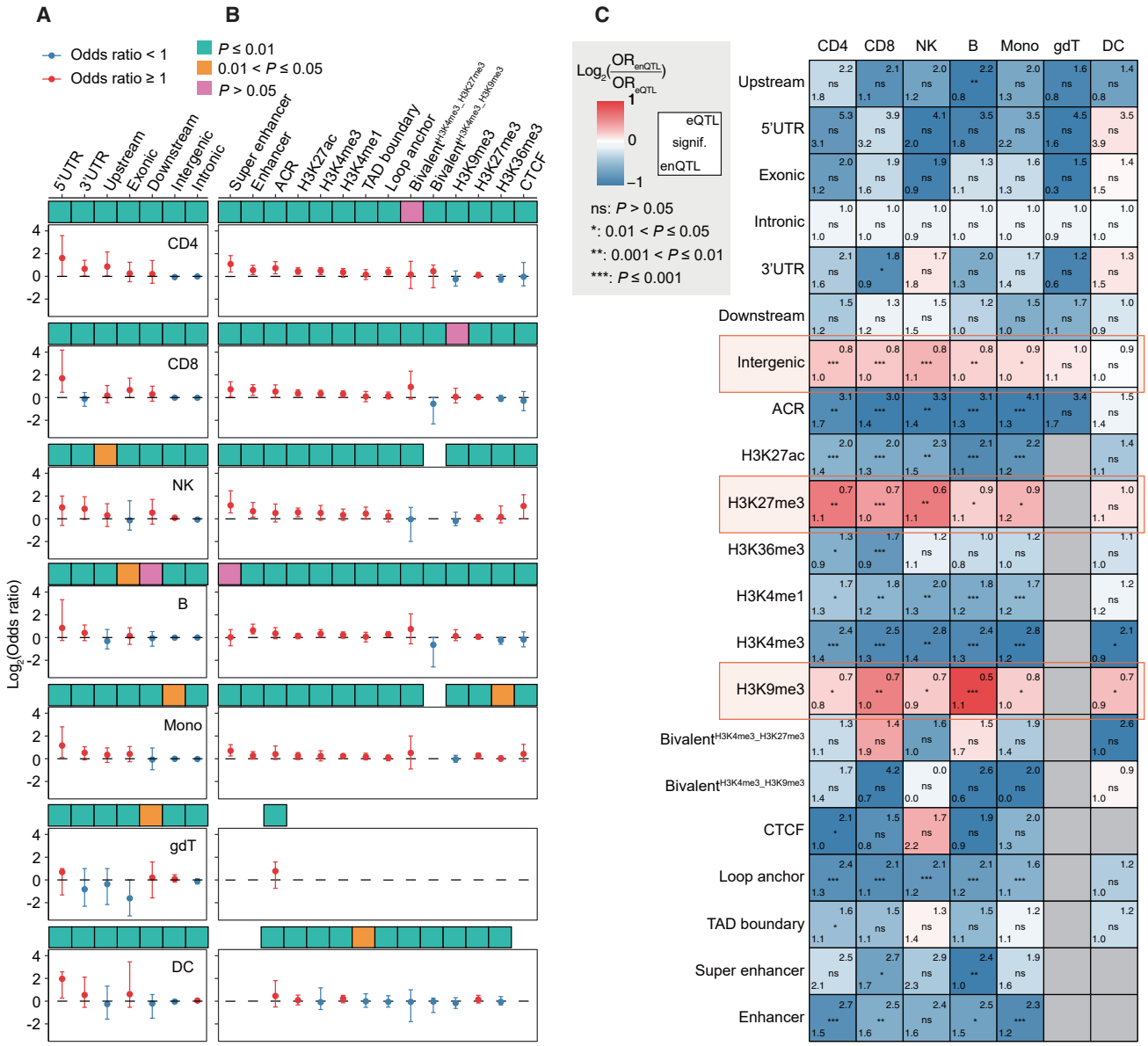


Figure 4. Potential regulatory mechanisms of enQTLs

(A and B) Odds ratios (ORs) depicting the enrichment of lead enQTL SNPs in various genomic regions (A) and epigenomic regions (B) relative to 1,000 permutation sets. Each column represents an estimated OR with error bars indicating the 95% confidence interval. The black dashed line marks the enrichment threshold (OR = 1). The cyan, orange, and pink squares represent $p \leq 0.01$, $0.01 < p \leq 0.05$, and $p > 0.05$, respectively. ORs and p values were calculated using Fisher's exact test. (C) ORs of genomic and epigenomic regions comparing enQTLs and eQTLs across cell types. The numbers in the lower left and upper right corners of each square represent the ORs of enQTLs and eQTLs, respectively. The significance in the center of the square indicates the comparison of enQTL and eQTL ORs. ns, $p > 0.05$; *$0.01 < p \leq 0.05$; **$0.001 < p \leq 0.01$; and ***$p \leq 0.001$. The color scale indicates the ratio of ORs between enQTLs and eQTLs: red signifies higher ORs for enQTLs, blue signifies higher ORs for eQTLs, and gray indicates missing annotation data for specific cell type.

and multiple diseases (colorectal cancer, hepatitis B virus infection, late-onset Alzheimer's disease, and age-related disorders).[43–46] In our data, rs1801133 is nominated from enQTL-colocalized events in CD8 cells targeting *KIAA2013*, and the cell-line-based chromatin immunoprecipitation sequencing (ChIP-seq) and Hi-C data supported the chromatin interaction between the promoters of *MTHFR* and *KIAA2013* (Figure 5B). Another case is the well-studied missense SNP, rs7412 (apolipoprotein E epsilon2) associated with various diseases (cardiovascular disease, hypertension, dementia) and human lifespan by regulating lipid metabolism.[47–51] In our data, rs7412 was nominated from enQTL-colocalized events in CD8 cells of long non-coding RNA (*CTB-129P6.11*), a target that has not previously been prioritized (Table S15).

There are several SNPs prioritized by colocalization that have been reported to be associated with other human traits. A

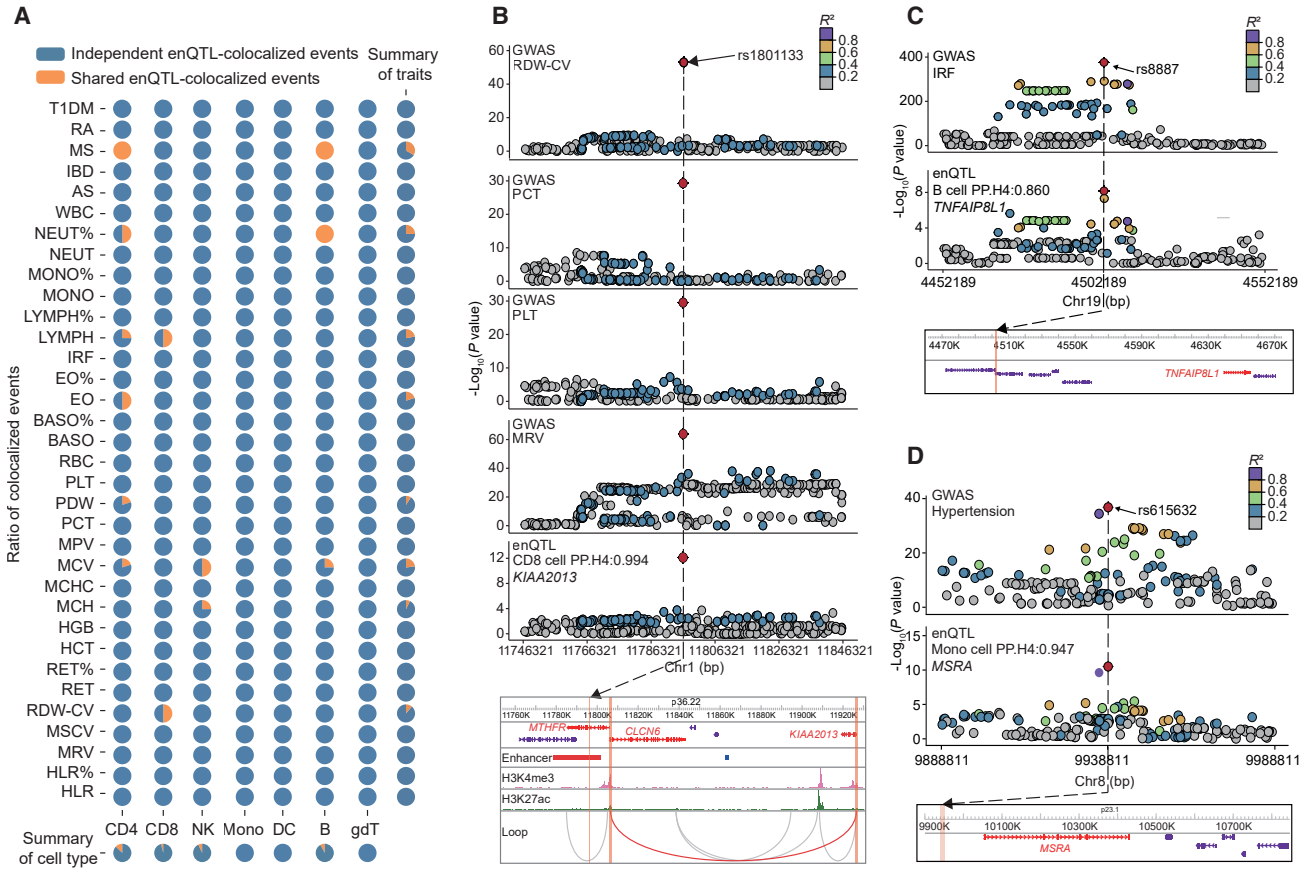


Figure 5. Colocalization between enQTL and GWAS signals of complex traits and diseases

(A) The pie chart presents the fraction of enQTL-colocalized events that are independent (blue) or shared (yellow) with eQTL-colocalized events across cell types and traits.

(B) The regional association plots of GWASs (RDW_CV, PCT, PLT, MRV), and enQTL of the gene *KIAA2013* within ±50 kb of rs1801133 is presented. Full-mode University of California, Santa Cruz (UCSC) gene tracks, as well as ChIP-seq peaks and Hi-C loops of CD8 cell line from ENCODE are displayed in this region. Three orange shades represent the location of rs1801133 and the left and right anchors of the Hi-C loops, respectively.

(C) The regional association plots of GWAS (IRF) and enQTL of the gene *TNFAIP8L1* within ±50 kb of rs8887 are presented.

(D) The regional association plots of GWAS (hypertension) and enQTL of the gene *MSRA* within ±50 kb of rs615632 are presented. MRV, mean reticulocyte volume; PCT, plateletcrit; PLT, platelet count; RDW_CV, red cell distribution width; NEUT%, neutrophil percentage; IFR, immature fraction of reticulocytes.

well-known obesity-associated SNP, rs8887,[52] was nominated by enQTL-colocalized events of target gene *TNFAIP8L1* with GWAS of five hematopoietic traits but no eQTL-colocalized events (Figure 5C; Table S15). An ankylosing spondylitis (AS)-associated SNP, rs11616188,[53–56] was nominated in the colocalization between GWAS traits of mean corpuscular hemoglobin (MCH) and mean corpuscular volume (MCV) and enQTL of *MLF2*, a Myeloid Leukemia Factor 2 implicated in leukemia (Figures S6C and S6D). Notably, rs140522 is another SNP prioritized from an enQTL-colocalized event between high light-scatter reticulocyte count (HLR) and target gene *TYMP* (Figure S6E). Multiple GWASs reported rs140522 to be significantly associated with leukemia.[57–59] Therefore, the role of *TYMP* as a dysregulated immunity gene in chronic lymphocytic leukemia could be hypothesized.[60]

To further investigate whether enQTLs contribute to other complex diseases, we extended our colocalization analysis to 1,932 diseases in the FinnGen cohort[61] and identified 411 enQTL-disease colocalization events in non-HLA regions (Table S17). A hypertension-associated SNP, rs615632, was nominated in the enQTL of *MSRA*, a gene related to lipid metabolism[62,63] (Figure 5D; Table S17).

Collectively, our results suggested that many enQTLs contribute to hematopoietic traits and autoimmune diseases, independent of eQTLs, thus providing cell-type-specific mechanistic hypotheses for these loci, including candidate causal variants, target genes, and cell types of action.

DISCUSSION

Understanding the genetic basis of complex traits and diseases is a fundamental question in human genetics and precision medicine. GWASs have identified numerous trait-associated loci and QTL mapping has been proved as a promising approach in interpreting their molecular mechanisms. However, the mainstream QTLs focus on steady-state average abundance of a trait (e.g.,

gene expression for eQTLs) and a considerable fraction of trait-associated loci lack eQTL support.[64] Alternatively, enQTL aims to detect loci that affect the variability in gene expression across individuals, which reflects underlying stochasticity or responses to subtle environmental fluctuations that are not captured in mean expression levels. In this study, we utilized population-scale single-cell data to establish the enQTL dataset across immune cell types in human peripheral blood. Our findings supported that most enQTLs are distinct from eQTLs and can explain a substantial proportion of previously unexplained loci, highlighting noise as an important mediator underlying genetic associations with complex traits such as hematopoietic traits and autoimmune diseases.

The transcriptome-wide noise atlas allows for two important observations. First, we observed a consistent pattern of elevated noise with age in six cell types. Previous studies have identified increased variability in the expression of immune response genes, attributing to destabilization of the immune-activation program in $CD4^+$ T cells.[12] Similarly, transcriptional variability was reported to increase with age in the human pancreas, along with increased stress signature and atypical hormone expression.[65] Here, we expanded the age-noise correlation as a consistent observation to multiple immune cell types, supporting the theory that molecular variability can be regarded as a signature of aging and is comparable across individuals as a quantitative trait. Second, we observed consistent pattern of elevated noise among males relative to female, independent of the aging effect. Previous studies reported an estrogen-dependent control of transcriptional burst in a breast cancer cell line[66] and *cis*-regulatory control of transcriptional noise in response to estrogen,[67] implying the role of hormone in noise regulation. In the evolutionary perspective, natural selection may favor the attenuated noise in female to ensure more consistent cellular function during critical periods (e.g., pregnancy), while males may potentially tolerate more noise in exchange for other fitness advantages (e.g., risk-taking behaviors).[68] In any case, it would be interesting to investigate the in-depth mechanisms of gender-specific noise pattern in future studies.

A key outcome of our work is to provide a set of enQTLs that could be contributing to human complex traits and diseases, which provides cell-type-specific mechanistic hypotheses for these loci, including candidate causal variants, target genes, and cell types of action. Noise is, intuitively, regarded as an undesired byproduct of biological systems and often conceives detrimental effects. Unexpectedly, our data supported that autoimmune-disease risk variants, in many cases, were associated with attenuated noise of nearby target genes. This unique directional effect implies the beneficial aspect of noise and highlights the evolving understanding of its double-edged role in evolutionary consequences. An interesting hypothesis is that the elevated immune response can be attributed to more sporadic transcriptional expression, such as a subpopulation of early-response cells with highly expressed immune-activation genes.[69] Specifically, *GNLY* is a lymphocyte-derived alarmin expressed by cytolytic T cells and NK cells within cytotoxic granules, serving dual functions in direct antimicrobial defense against intracellular pathogens (e.g., *Mycobacterium tuberculosis*) and indirect immune activation through recruitment and stimulation of antigen-presenting cells.[19,70] The *GNLY* expression noise may tune population-level immune responses through stochastic threshold effects. Individuals with early-response cells were thus characterized with elevated noise and showed higher response strength in the immune system.

The biological mechanisms underlying enQTL effects may operate across multiple levels. At the DNA level, variants located in promoters or enhancers can alter TF-binding affinity or CpG-island stability, with stochastic TF binding acting as a source of expression noise.[71] At the epigenetic level, variants may influence DNA methylation or histone modifications, thereby modulating chromatin accessibility and nucleosome positioning. These changes can regulate transcriptional bursting dynamics and amplify cell-to-cell heterogeneity.[72] At the transcriptional level, enQTLs may affect RNA polymerase II recruitment, initiation kinetics, or pause-release dynamics, tuning transcriptional stochasticity without necessarily altering mean expression levels. Repressive *cis*-regulatory elements—such as those mediated by polycomb complexes or DNA methylation—function as dynamic dampers of transcriptional stochasticity. As reported,[72] these elements modulate noise through three interconnected mechanisms: stabilizing chromatin states to limit spontaneous transcriptional bursts, enforcing spatiotemporal precision to reduce cell-to-cell variability, and suppressing random fluctuations to enhance expression robustness. Our finding that enQTLs are significantly enriched in repressive *cis*-regulatory regions suggests that genetic variants in these contexts may fine-tune transcriptional noise levels by modulating repression dynamics. For example, a noise-reducing allele could strengthen polycomb binding or slow chromatin state transitions, whereas a noise-increasing allele might weaken repressor occupancy or accelerate stochastic switching.

We have instead implemented multiple safeguards to mitigate potential biases from single-cell RNA sequencing (scRNA-seq) dropouts, particularly for low- to medium-expressed genes, including stringent expression filtering (mean unique molecular identifier >0.1), gamma distribution modeling of mean-variability dependence to account for over-dispersion,[3] and variance stabilization via SCTransform normalization. These statistical approaches are well established in the field and have been applied successfully in other studies without experimental validation.[22,73–75] Furthermore, our noise estimates showed robust replication in an independent 5′ 10x scRNA-seq cohort of 149 Japanese individuals[34] and are consistent with previous work demonstrating close agreement between statistically derived and single-molecule fluorescence *in situ* hybridization (smFISH)-measured coefficients of variation, even for low-expression genes such as *Gli2* and *Stag3*.[75] Together, these considerations suggest that, while experimental validation remains an important future direction, the statistical framework applied here provides a reliable basis for identifying genetic determinants of transcriptional noise.

It should also be noted that very few of the enQTL-colocalizing genetic loci overlapped with eQTLs, indicating that enQTLs are largely independent. This limited overlap between enQTLs and eQTLs suggests mechanisms for tuning expression noise

independently of mean output. Two experimentally established pathways provide explanations. First, promoter-centered kinetic tuning[71] demonstrates how subtle sequence variants in core regulatory elements (e.g., 5′ UTRs/promoters) can selectively alter transcriptional burst size—a primary determinant of noise—while preserving burst frequency that governs mean expression. This occurs through precision adjustments in RNAPII initiation efficiency or TF binding cooperativity, directly aligning with our observed enQTL enrichment in these regions. Second, DNA-repair-mediated stochasticity[76] offers an orthogonal mechanism where variants at topological "noise hotspots" (e.g., APEX1-bound sites) modulate local supercoiling dynamics. By altering the recruitment of DNA-repair complexes, such variants generate stochastic torsional stress that amplifies transcriptional noise without affecting basal expression levels. Notably, DNA repair genes *APEX1* and *TK1*—previously validated to modulate expression noise via CRISPRi perturbation—were additionally identified as enGenes in B cells in our study. In addition, other distinct biological processes and regulatory mechanisms also could be the reasons for this few overlap, such as post-transcriptional control (e.g., mRNA stability)[77] and epigenetic dynamics (e.g., bivalent promoters).[7] It could be anticipated that this enQTL dataset with full summary statistics could be helpful for future studies to identify functional genes and variants for other complex traits.

Collectively, our study provides comprehensive insights into the genetic determinants of gene expression noise in the human genome and a valuable resource for understanding the role of expression noise in human complex traits and disease susceptibility. Our findings underscore the importance of noise as a biologically meaningful molecular trait; however, future work incorporating detailed mechanistic experiments will be essential to interpret the regulation of noise.

Limitations of the study

A principal limitation of this study is the absence of orthogonal experimental validation (e.g., RNA fluorescence *in situ* hybridization [FISH]) for the large set of 6,743 enGenes identified across seven immune cell types. While such assays are valuable for corroborating biological noise estimates, current RNA FISH technologies are low throughput (typically interrogating fewer than 10 genes per study) and are not readily compatible with the large-scale primary human cohort samples analyzed here. Emerging multiplexed FISH platforms may help address this limitation by enabling higher-throughput validation in future work.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Erping Long (erping.long@ibms.pumc.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- GWAS summary statistics of gene expression noise across seven immune cell types are available at Figshare (https://figshare.com/s/ea4eca9784af800db051). The following publicly available datasets were used in this study: FinnGen, https://www.finngen.fi/en/access_results; GWAS summary statistics of blood traits, ftp://ftp.sanger.ac.uk/pub/project/humgen/summary_statistics/UKBB_blood_cell_traits; IBD, https://ftp.sanger.ac.uk/pub/project/humgen/summary_statistics/human/2016-11-07; SLE, http://urr.cat/data/GWAS_SLE_summaryStats.zip; MS, http://imsgc.net/data/imsgc_mssev_discovery.tar.gz. The GWAS Catalog accession numbers for summary statistics: RA, GCST90132001-GCST90133000; AS, GCST005001-GCST006000; T1DM, GCST90014001-GCST90015000; AD, GCST90012001-GCST90013000. The scRNA-seq data of the OneK1K cohort: https://cellxgene.cziscience.com/collections/dde06e0f-ab3b-46be-96a2-a8082383c4a1. The scRNA-seq data of 149 Japanese individuals from AIDA cohort: https://explore.data.humancellatlas.org/projects/35d5b057-3daf-4ccd-8112-196194598893. OpenTarget Genetics: https://genetics.opentargets.org.
- All original code has been deposited at Zenodo at https://doi.org/10.5281/zenodo.17239294 and is publicly available as of the date of publication.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

E.L. is supported by the National Natural Science Foundation of China (Excellent Youth Scholars Program, 32300483, 32470635, and 82090011), National Key R&D Program (2024ZD0528500), State Key Laboratory Special Fund (2060204), and Chinese Academy of Medical Sciences Innovation Fund (2023-I2M-3-010, 2023-I2M-2-001). We thank Xushen Xiong, Jianzhi Zhang, Zhiyue Zhang, Choi lab members, and Long lab members for valuable comments. We thank the Bioinformatics Center of Institute of Basic Medical Sciences and Biomedical High Performance Computing Platform, Chinese Academy of Medical Sciences for computing support, and Furen Shi and Lifang Xie for technical assistance.

AUTHOR CONTRIBUTIONS

E.L., Y.L., and T.Z. contributed to the conception and design of the work. Y.L., X.N., T.C., X.H., Q.H., J.W., and C.W. contributed to the data acquisition, curation, and analysis. Z.H., H.X., M.S., J.P., and J.C. provided valuable assistance. All authors contributed to the drafting and revising of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the authors used ChatGPT-4 and DeepSeek-R1 to ensure the grammar, clarity, precision, and impact of the writing. After using this tool or service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- METHOD DETAILS
  - Quality control of the genotype data
  - Imputation of genotyping data
  - Expression data processing
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Gene expression noise calculation
  - HVG and individual-level noise calculation
  - enQTL identification
  - Latent factor identification and removal

- enQTL validation
- Cell-type-specific correlations between TF and enGene
- Functional annotation and enrichment analysis of enQTL
- Colocalization analysis between enQTL and GWAS loci

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2025.116612.

Received: December 10, 2024
Revised: August 16, 2025
Accepted: November 4, 2025

REFERENCES

1. Eling, N., Morgan, M.D., and Marioni, J.C. (2019). Challenges in measuring and understanding biological noise. Nat. Rev. Genet. *20*, 536–548. https://doi.org/10.1038/s41576-019-0130-6.
2. Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic Gene Expression in a Single Cell. Science *297*, 1183–1186. https://doi.org/10.1126/science.1070919.
3. Brennecke, P., Anders, S., Kim, J.K., Kołodziejczyk, A.A., Zhang, X., Proserpio, V., Baying, B., Benes, V., Teichmann, S.A., Marioni, J.C., and Heisler, M.G. (2013). Accounting for technical noise in single-cell RNA-seq experiments. Nat. Methods *10*, 1093–1095. https://doi.org/10.1038/nmeth.2645.
4. Tirosh, I., Weinberger, A., Carmi, M., and Barkai, N. (2006). A genetic signature of interspecies variations in gene expression. Nat. Genet. *38*, 830–834. https://doi.org/10.1038/ng1819.
5. Tirosh, I., and Barkai, N. (2008). Two strategies for gene regulation by promoter nucleosomes. Genome Res. *18*, 1084–1091. https://doi.org/10.1101/gr.076059.108.
6. Zoller, B., Nicolas, D., Molina, N., and Naef, F. (2015). Structure of silent transcription intervals and noise characteristics of mammalian genes. Mol. Syst. Biol. *11*, 823. https://doi.org/10.15252/msb.20156257.
7. Faure, A.J., Schmiedel, J.M., and Lehner, B. (2017). Systematic Analysis of the Determinants of Gene Expression Noise in Embryonic Stem Cells. Cell Syst. *5*, 471–484.e4. https://doi.org/10.1016/j.cels.2017.10.003.
8. Morgan, M.D., and Marioni, J.C. (2018). CpG island composition differences are a source of gene expression noise indicative of promoter responsiveness. Genome Biol. *19*, 81. https://doi.org/10.1186/s13059-018-1461-x.
9. Goolam, M., Scialdone, A., Graham, S.J.L., Macaulay, I.C., Jedrusik, A., Hupalowska, A., Voet, T., Marioni, J.C., and Zernicka-Goetz, M. (2016). Heterogeneity in Oct4 and Sox2 Targets Biases Cell Fate in 4-Cell Mouse Embryos. Cell *165*, 61–74. https://doi.org/10.1016/j.cell.2016.01.047.
10. Mohammed, H., Hernando-Herraez, I., Savino, A., Scialdone, A., Macaulay, I., Mulas, C., Chandra, T., Voet, T., Dean, W., Nichols, J., et al. (2017). Single-Cell Landscape of Transcriptional Heterogeneity and Cell Fate Decisions during Mouse Early Gastrulation. Cell Rep. *20*, 1215–1228. https://doi.org/10.1016/j.celrep.2017.07.009.
11. Ohnishi, Y., Huber, W., Tsumura, A., Kang, M., Xenopoulos, P., Kurimoto, K., Oleś, A.K., Araúzo-Bravo, M.J., Saitou, M., Hadjantonakis, A.-K., and Hiiragi, T. (2014). Cell-to-cell expression variability followed by signal reinforcement progressively segregates early mouse lineages. Nat. Cell Biol. *16*, 27–37. https://doi.org/10.1038/ncb2881.
12. Martinez-Jimenez, C.P., Eling, N., Chen, H.-C., Vallejos, C.A., Kolodziejczyk, A.A., Connor, F., Stojic, L., Rayner, T.F., Stubbington, M.J.T., Teichmann, S.A., et al. (2017). Aging increases cell-to-cell transcriptional variability upon immune stimulation. Science *355*, 1433–1436. https://doi.org/10.1126/science.aah4115.
13. Sun, M., and Zhang, J. (2020). Allele-specific single-cell RNA sequencing reveals different architectures of intrinsic and extrinsic gene expression noises. Nucleic Acids Res. *48*, 533–547. https://doi.org/10.1093/nar/gkz1134.
14. Abdellaoui, A., Yengo, L., Verweij, K.J.H., and Visscher, P.M. (2023). 15 years of GWAS discovery: Realizing the promise. Am. J. Hum. Genet. *110*, 179–194. https://doi.org/10.1016/j.ajhg.2022.12.011.
15. Gallagher, M.D., and Chen-Plotkin, A.S. (2018). The Post-GWAS Era: From Association to Function. Am. J. Hum. Genet. *102*, 717–730. https://doi.org/10.1016/j.ajhg.2018.04.002.
16. Yazar, S., Alquicira-Hernandez, J., Wing, K., Senabouth, A., Gordon, M.G., Andersen, S., Lu, Q., Rowson, A., Taylor, T.R.P., Clarke, L., et al. (2022). Single-cell eQTL mapping identifies cell type–specific genetic control of autoimmune disease. Science *376*, eabf3041. https://doi.org/10.1126/science.abf3041.
17. Dey, S.S., Foley, J.E., Limsirichai, P., Schaffer, D.V., and Arkin, A.P. (2015). Orthogonal control of expression mean and variance by epigenetic features at different genomic loci. Mol. Syst. Biol. *11*, 806. https://doi.org/10.15252/msb.20145704.
18. Mandric, I., Schwarz, T., Majumdar, A., Hou, K., Briscoe, L., Perez, R., Subramaniam, M., Hafemeister, C., Satija, R., Ye, C.J., et al. (2020). Optimized design of single-cell RNA sequencing experiments for cell-type-specific eQTL analysis. Nat. Commun. *11*, 5504. https://doi.org/10.1038/s41467-020-19365-w.
19. Tewary, P., Yang, D., De La Rosa, G., Li, Y., Finn, M.W., Krensky, A.M., Clayberger, C., and Oppenheim, J.J. (2010). Granulysin activates antigen-presenting cells through TLR4 and acts as an immune alarmin. Blood *116*, 3465–3474. https://doi.org/10.1182/blood-2010-03-273953.
20. Lanningham-Foster, L., Green, C.L., Langkamp-Henken, B., Davis, B.A., Nguyen, K.T., Bender, B.S., and Cousins, R.J. (2002). Overexpression of CRIP in transgenic mice alters cytokine patterns and the immune response. Am. J. Physiol. Endocrinol. Metab. *282*, E1197–E1203. https://doi.org/10.1152/ajpendo.00508.2001.
21. Latonen, L., Järvinen, P.M., and Laiho, M. (2008). Cytoskeleton-interacting LIM-domain protein CRP1 suppresses cell proliferation and protects from stress-induced cell death. Exp. Cell Res. *314*, 738–747. https://doi.org/10.1016/j.yexcr.2007.11.024.
22. Osorio, D., Yu, X., Zhong, Y., Li, G., Yu, P., Serpedin, E., Huang, J.Z., and Cai, J.J. (2019). Single-Cell Expression Variability Implies Cell Function. Cells *9*, 14. https://doi.org/10.3390/cells9010014.
23. Zheng, H., Vijg, J., Fard, A.T., and Mar, J.C. (2023). Measuring cell-to-cell expression variability in single-cell RNA-sequencing data: a comparative analysis and applications to B cell aging. Genome Biol. *24*, 238. https://doi.org/10.1186/s13059-023-03036-2.
24. Angelidis, I., Simon, L.M., Fernandez, I.E., Strunz, M., Mayr, C.H., Greiffo, F.R., Tsitsiridis, G., Ansari, M., Graf, E., Strom, T.-M., et al. (2019). An atlas of the aging lung mapped by single cell transcriptomics and deep tissue proteomics. Nat. Commun. *10*, 963. https://doi.org/10.1038/s41467-019-08831-9.
25. Lu, Y., Biancotto, A., Cheung, F., Remmers, E., Shah, N., McCoy, J.P., and Tsang, J.S. (2016). Systematic Analysis of Cell-to-Cell Expression Variation of T Lymphocytes in a Human Cohort Identifies Aging and Genetic Associations. Immunity *45*, 1162–1175. https://doi.org/10.1016/j.immuni.2016.10.025.
26. Ugel, S., Canè, S., De Sanctis, F., and Bronte, V. (2021). Monocytes in the Tumor Microenvironment. Annu. Rev. Pathol. Mech. Dis. *16*, 93–122. https://doi.org/10.1146/annurev-pathmechdis-012418-013058.
27. Taylor-Weiner, A., Aguet, F., Haradhvala, N.J., Gosai, S., Anand, S., Kim, J., Ardlie, K., Van Allen, E.M., and Getz, G. (2019). Scaling computational genomics to millions of individuals with GPUs. Genome Biol. *20*, 228. https://doi.org/10.1186/s13059-019-1836-7.
28. Stegle, O., Parts, L., Durbin, R., and Winn, J. (2010). A Bayesian Framework to Account for Complex Non-Genetic Factors in Gene Expression

Levels Greatly Increases Power in eQTL Studies. PLoS Comput. Biol. *6*, e1000770. https://doi.org/10.1371/journal.pcbi.1000770.

29. Xiong, X., James, B.T., Boix, C.A., Park, Y.P., Galani, K., Victor, M.B., Sun, N., Hou, L., Ho, L.-L., Mantero, J., et al. (2023). Epigenomic dissection of Alzheimer's disease pinpoints causal variants and reveals epigenome erosion. Cell *186*, 4422–4437.e21. https://doi.org/10.1016/j.cell.2023.08.040.
30. Xiong, X., Hou, L., Park, Y.P., Molinie, B., GTEx Consortium; Gregory, R.I., Kellis, M., Gregory, R.I., and Kellis, M. (2021). Genetic drivers of m6A methylation in human brain, lung, heart and muscle. Nat. Genet. *53*, 1156–1165. https://doi.org/10.1038/s41588-021-00890-3.
31. Cheng, C.-W., Fang, W.-F., and Lin, J.-D. (2023). Associations of serum keratin 1 with thyroid function and immunity in Graves' disease. PLoS One *18*, e0289345. https://doi.org/10.1371/journal.pone.0289345.
32. Khalid, H.N., Elghobashy, Y.A.E., and Elsayed, A.N. (2022). GNLY gene polymorphism: A potential role in understanding psoriasis pathogenesis. J. Cosmet. Dermatol. *21*, 4805–4809. https://doi.org/10.1111/jocd.14792.
33. Fernandez-Fernandez, M.R., Veprintsev, D.B., and Fersht, A.R. (2005). Proteins of the S100 family regulate the oligomerization of p53 tumor suppressor. Proc. Natl. Acad. Sci. USA *102*, 4735–4740. https://doi.org/10.1073/pnas.0501459102.
34. Kock, K.H., Tan, L.M., Han, K.Y., Ando, Y., Jevapatarakul, D., Chatterjee, A., Lin, Q.X.X., Buyamin, E.V., Sonthalia, R., Rajagopalan, D., et al. (2025). Asian diversity in human immune cells. Cell *188*, 2288–2306.e24. https://doi.org/10.1016/j.cell.2025.02.017.
35. Khakoo, S.I., Thio, C.L., Martin, M.P., Brooks, C.R., Gao, X., Astemborski, J., Cheng, J., Goedert, J.J., Vlahov, D., Hilgartner, M., et al. (2004). HLA and NK Cell Inhibitory Receptor Genes in Resolving Hepatitis C Virus Infection. Science *305*, 872–874. https://doi.org/10.1126/science.1097670.
36. Rezaieyazdi, Z., Tavakkol-Afshari, J., Esmaili, E., Orouji, E., Pezeshkpour, F., Khodadoost, M., Mazhani, M., and Sandooghi, M. (2008). Association of HLA-DQB1 allelic sequence variation with susceptibility to systemic lupus erythematosus. Iran. J. Allergy Asthma Immunol. *7*, 91–95.
37. Strnadel, J., Choi, S., Fujimura, K., Wang, H., Zhang, W., Wyse, M., Wright, T., Gross, E., Peinado, C., Park, H.W., et al. (2017). eIF5A-PEAK1 Signaling Regulates YAP1/TAZ Protein Expression and Pancreatic Cancer Cell Growth. Cancer Res. *77*, 1997–2007. https://doi.org/10.1158/0008-5472.CAN-16-2594.
38. Vaughan, N., Scholz, N., Lindon, C., and Licchesi, J.D.F. (2022). The E3 ubiquitin ligase HECTD1 contributes to cell proliferation through an effect on mitosis. Sci. Rep. *12*, 13160. https://doi.org/10.1038/s41598-022-16965-y.
39. Urbut, S.M., Wang, G., Carbonetto, P., and Stephens, M. (2019). Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. Nat. Genet. *51*, 187–195. https://doi.org/10.1038/s41588-018-0268-8.
40. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. *38*, e164. https://doi.org/10.1093/nar/gkq603.
41. ENCODE Project Consortium (2004). The ENCODE (ENCyclopedia Of DNA Elements) Project. Science *306*, 636–640. https://doi.org/10.1126/science.1105136.
42. Wang, Y., Song, C., Zhao, J., Zhang, Y., Zhao, X., Feng, C., Zhang, G., Zhu, J., Wang, F., Qian, F., et al. (2023). SEdb 2.0: a comprehensive super-enhancer database of human and mouse. Nucleic Acids Res. *51*, D280–D290. https://doi.org/10.1093/nar/gkac968.
43. Paré, G., Chasman, D.I., Parker, A.N., Zee, R.R.Y., Mälarstig, A., Seedorf, U., Collins, R., Watkins, H., Hamsten, A., Miletich, J.P., and Ridker, P.M. (2009). Novel Associations of CPS1, MUT, NOX4, and DPEP1 With Plasma Homocysteine in a Healthy Population: A Genome-Wide Evaluation of 13 974 Participants in the Women's Genome Health Study. Circ. Cardiovasc. Genet. *2*, 142–150. https://doi.org/10.1161/CIRCGENETICS.108.829804.
44. Teng, Z., Wang, L., Cai, S., Yu, P., Wang, J., Gong, J., and Liu, Y. (2013). The 677C>T (rs1801133) Polymorphism in the MTHFR Gene Contributes to Colorectal Cancer Risk: A Meta-Analysis Based on 71 Research Studies. PLoS One *8*, e55332. https://doi.org/10.1371/journal.pone.0055332.
45. Karakus, N., Yigit, S., Kocak, M., Bozkurt, N., and Duygu, F. (2023). Impact of methylene-tetrahydrofolate reductase gene C677T and A1298C polymorphisms as a risk factor for hepatitis B virus infection. Nucleosides Nucleotides Nucleic Acids *42*, 683–695. https://doi.org/10.1080/15257770.2023.2187061.
46. Sasner, M., Preuss, C., Pandey, R.S., Uyar, A., Garceau, D., Kotredes, K.P., Williams, H., Oblak, A.L., Lin, P.B.C., Perkins, B., et al. (2024). In vivo validation of late-onset Alzheimer's disease genetic risk factors. Alzheimer's Dement. *20*, 4970–4984. https://doi.org/10.1002/alz.13840.
47. Deelen, J., Evans, D.S., Arking, D.E., Tesi, N., Nygaard, M., Liu, X., Wojczynski, M.K., Biggs, M.L., Van Der Spek, A., Atzmon, G., et al. (2019). A meta-analysis of genome-wide association studies identifies multiple longevity genes. Nat. Commun. *10*, 3669. https://doi.org/10.1038/s41467-019-11558-2.
48. Katz, D.H., Tahir, U.A., Bick, A.G., Pampana, A., Ngo, D., Benson, M.D., Yu, Z., Robbins, J.M., Chen, Z.-Z., Cruz, D.E., et al. (2022). Whole Genome Sequence Analysis of the Plasma Proteome in Black Adults Provides Novel Insights Into Cardiovascular Disease. Circulation *145*, 357–370. https://doi.org/10.1161/CIRCULATIONAHA.121.055117.
49. Wang, C., Ji, X., Tang, Z., Zhang, Z., Gu, X., and Fang, X. (2022). Combined homocysteine and apoE rs429358 and rs7412 polymorphism in association with serum lipid levels and cognition in Chinese community-dwelling older adults. BMC Psychiatry *22*, 223. https://doi.org/10.1186/s12888-022-03877-4.
50. Rao, H., Wu, H., Yu, Z., and Huang, Q. (2022). APOE Genetic Polymorphism rs7412 T/T Genotype May Be a Risk Factor for Essential Hypertension among Hakka People in Southern China. Int. J. Hypertens. *2022*, 8145896–8145899. https://doi.org/10.1155/2022/8145896.
51. Zhou, C., Zhang, Y., Yang, S., Ye, Z., Wu, Q., Liu, M., He, P., Zhang, Y., and Qin, X. (2023). Habitual glucosamine use, APOE genotypes, and risk of incident cause-specific dementia in the older population. Alz Res Therapy *15*, 152. https://doi.org/10.1186/s13195-023-01295-6.
52. Richardson, K., Louie-Gao, Q., Arnett, D.K., Parnell, L.D., Lai, C.-Q., Davalos, A., Fox, C.S., Demissie, S., Cupples, L.A., Fernandez-Hernando, C., and Ordovas, J.M. (2011). The PLIN4 Variant rs8887 Modulates Obesity Related Phenotypes in Humans through Creation of a Novel miR-522 Seed Site. PLoS One *6*, e17944. https://doi.org/10.1371/journal.pone.0017944.
53. Evans, D.M., Spencer, C.C.A., Pointon, J.J., Su, Z., Harvey, D., Kochan, G., Oppermann, U., Dilthey, A., Pirinen, M., Stone, M.A., et al. (2011). Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. Nat. Genet. *43*, 761–767. https://doi.org/10.1038/ng.873.
54. Wei, J.C.-C., Hsu, Y.-W., Hung, K.-S., Wong, R.-H., Huang, C.-H., Liu, Y.-T., Guo, Y.-C., Ikegawa, S., and Chang, W.-C. (2013). Association Study of Polymorphisms rs4552569 and rs17095830 and the Risk of Ankylosing Spondylitis in a Taiwanese Population. PLoS One *8*, e52801. https://doi.org/10.1371/journal.pone.0052801.
55. Wen, Y.-F., Wei, J.C.-C., Hsu, Y.-W., Chiou, H.-Y., Wong, H.S.-C., Wong, R.-H., Ikegawa, S., and Chang, W.-C. (2014). rs10865331 Associated with Susceptibility and Disease Severity of Ankylosing Spondylitis in a Taiwanese Population. PLoS One *9*, e104525. https://doi.org/10.1371/journal.pone.0104525.
56. Jia, H., Chen, M., Cai, Y., Luo, X., Hou, G., Li, Y., Cai, C., Chen, J., Li, Q., Miu, K.-K., et al. (2022). A new and spontaneous animal model for ankylosing spondylitis is found in cynomolgus monkeys. Arthritis Res. Ther. *24*, 1. https://doi.org/10.1186/s13075-021-02679-5.
57. Jeong, S., Tsai, M.-J., Shen, C., and Hsu, Y.-H. (2024). Falls, fracture and frailty risk in multiple sclerosis: a Mendelian Randomization study to

identify shared genetics. J. Bone Miner. Metab. *42*, 335–343. https://doi.org/10.1007/s00774-024-01504-8.

58. Ryu, J., Woo, J., Shin, J., Ryoo, H., Kim, Y., and Lee, C. (2014). Profile of Differential Promoter Activity by Nucleotide Substitution at GWAS Signals For Multiple Sclerosis. Medicine (Baltim.) *93*, e281. https://doi.org/10.1097/MD.0000000000000281.
59. Giallourakis, C.C., Benita, Y., Molinie, B., Cao, Z., Despo, O., Pratt, H.E., Zukerberg, L.R., Daly, M.J., Rioux, J.D., and Xavier, R.J. (2013). Genome-wide Analysis of Immune System Genes by Expressed Sequence Tag Profiling. J. Immunol. *190*, 5578–5587. https://doi.org/10.4049/jimmunol.1203471.
60. Law, P.J., Berndt, S.I., Speedy, H.E., Camp, N.J., Sava, G.P., Skibola, C.F., Holroyd, A., Joseph, V., Sunter, N.J., Nieters, A., et al. (2017). Genome-wide association analysis implicates dysregulation of immunity genes in chronic lymphocytic leukaemia. Nat. Commun. *8*, 14175. https://doi.org/10.1038/ncomms14175.
61. Kurki, M.I., Karjalainen, J., Palta, P., Sipilä, T.P., Kristiansson, K., Donner, K.M., Reeve, M.P., Laivuori, H., Aavikko, M., Kaunisto, M.A., et al. (2023). FinnGen provides genetic insights from a well-phenotyped isolated population. Nature *613*, 508–518. https://doi.org/10.1038/s41586-022-05473-8.
62. Kazawa, T., Kawasaki, T., Sakamoto, A., Imamura, M., Ohashi, R., Jiang, S., Tanaka, T., Iwanari, H., Hamakubo, T., Sakai, J., et al. (2009). Expression of liver X receptor alpha and lipid metabolism in granulocyte-macrophage colony-stimulating factor-induced human monocyte-derived macrophage. Pathol Int. *59*, 152.
63. Sigalov, A.B., and Stern, L.J. (1998). Enzymatic repair of oxidative damage to human apolipoprotein A-I. FEBS Lett. *433*, 196–200. https://doi.org/10.1016/S0014-5793(98)00908-9.
64. The, G.T.E.C., Aguet, F., Anand, S., Ardlie, K.G., Gabriel, S., Getz, G.A., Graubert, A., Hadley, K., Handsaker, R.E., Huang, K.H., et al. (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science *369*, 1318–1330. https://doi.org/10.1126/science.aaz1776.
65. Enge, M., Arda, H.E., Mignardi, M., Beausang, J., Bottino, R., Kim, S.K., and Quake, S.R. (2017). Single-Cell Analysis of Human Pancreas Reveals Transcriptional Signatures of Aging and Somatic Mutation Patterns. Cell *171*, 321–330.e14. https://doi.org/10.1016/j.cell.2017.09.004.
66. Fritzsch, C., Baumgärtner, S., Kuban, M., Steinshorn, D., Reid, G., and Legewie, S. (2018). Estrogen-dependent control and cell-to-cell variability of transcriptional bursting. Mol. Syst. Biol. *14*, e7678. https://doi.org/10.15252/msb.20177678.
67. Ginley-Hidinger, M., Abewe, H., Osborne, K., Richey, A., Kitchen, N., Mortenson, K.L., Wissink, E.M., Lis, J., Zhang, X., and Gertz, J. (2024). Cis-regulatory control of transcriptional timing and noise in response to estrogen. Cell Genom. *4*, 100542. https://doi.org/10.1016/j.xgen.2024.100542.
68. Song, S., and Zhang, J. (2024). Genetic variants underlying human bisexual behavior are reproductively advantageous. Sci. Adv. *10*, eadj6958. https://doi.org/10.1126/sciadv.adj6958.
69. Goronzy, J.J., and Weyand, C.M. (2013). Understanding immunosenescence to improve responses to vaccines. Nat. Immunol. *14*, 428–436. https://doi.org/10.1038/ni.2588.
70. Stenger, S., Hanson, D.A., Teitelbaum, R., Dewan, P., Niazi, K.R., Froelich, C.J., Ganz, T., Thoma-Uszynski, S., Melián, A., Bogdan, C., et al. (1998). An Antimicrobial Activity of Cytolytic T Cells Mediated by Granulysin. Science *282*, 121–125. https://doi.org/10.1126/science.282.5386.121.
71. Sharon, E., van Dijk, D., Kalma, Y., Keren, L., Manor, O., Yakhini, Z., and Segal, E. (2014). Probing the effect of promoters on noise in gene expression using thousands of designed sequences. Genome Res. *24*, 1698–1706. https://doi.org/10.1101/gr.168773.113.
72. Cassidy, J.J., Bernasek, S.M., Bakker, R., Giri, R., Peláez, N., Eder, B., Bobrowska, A., Bagheri, N., Nunes Amaral, L.A., and Carthew, R.W. (2019). Repressive Gene Regulation Synchronizes Development with Cellular Metabolism. Cell *178*, 980–992.e17. https://doi.org/10.1016/j.cell.2019.06.023.
73. Vallejos, C.A., Marioni, J.C., and Richardson, S. (2015). BASiCS: Bayesian Analysis of Single-Cell Sequencing Data. PLoS Comput. Biol. *11*, e1004333. https://doi.org/10.1371/journal.pcbi.1004333.
74. Eling, N., Richard, A.C., Richardson, S., Marioni, J.C., and Vallejos, C.A. (2018). Correcting the Mean-Variance Dependency for Differential Variability Testing Using Single-Cell RNA Sequencing Data. Cell Syst. *7*, 284–294.e12. https://doi.org/10.1016/j.cels.2018.06.011.
75. Grün, D., Kester, L., and van Oudenaarden, A. (2014). Validation of noise models for single-cell transcriptomics. Nat. Methods *11*, 637–640. https://doi.org/10.1038/nmeth.2930.
76. Desai, R.V., Chen, X., Martin, B., Chaturvedi, S., Hwang, D.W., Li, W., Yu, C., Ding, S., Thomson, M., Singer, R.H., et al. (2021). A DNA repair pathway can regulate transcriptional noise to promote cell fate transitions. Science *373*, eabc6506. https://doi.org/10.1126/science.abc6506.
77. Baudrimont, A., Jaquet, V., Wallerich, S., Voegeli, S., and Becskei, A. (2019). Contribution of RNA Degradation to Intrinsic and Extrinsic Noise in Gene Expression. Cell Rep. *26*, 3752–3761.e5. https://doi.org/10.1016/j.celrep.2019.03.001.
78. Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew, E.Y., Levy, S., McGue, M., et al. (2016). Next-generation genotype imputation service and methods. Nat. Genet. *48*, 1284–1287. https://doi.org/10.1038/ng.3656.
79. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., De Bakker, P.I.W., Daly, M.J., and Sham, P.C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am. J. Hum. Genet. *81*, 559–575. https://doi.org/10.1086/519795.
80. Wang, G., Sarkar, A., Carbonetto, P., and Stephens, M. (2020). A Simple New Approach to Variable Selection in Regression, with Application to Genetic Fine Mapping. J. R. Stat. Soc. Series B Stat. Methodol. *82*, 1273–1300. https://doi.org/10.1111/rssb.12388.
81. Coetzee, S.G., and Hazelett, D.J. (2024). motifbreakR v2: expanded variant analysis including indels and integrated evidence from transcription factor binding databases. Bioinform. Adv. *4*, vbae162. https://doi.org/10.1093/bioadv/vbae162.
82. Hafemeister, C., and Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol. *20*, 296. https://doi.org/10.1186/s13059-019-1874-1.
83. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
84. Aguiar, V.R.C., César, J., Delaneau, O., Dermitzakis, E.T., and Meyer, D. (2019). Expression estimation and eQTL mapping for HLA genes with a personalized pipeline. PLoS Genet. *15*, e1008091. https://doi.org/10.1371/journal.pgen.1008091.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
GWAS summary statistics of gene expression noise	This paper.	Figshare: https://figshare.com/s/ea4eca9784af800db051
OneK1K cohort	Yazar et al.[16]	https://cellxgene.cziscience.com/collections/dde06e0f-ab3b-46be-96a2-a8082383c4a1
Japanese individuals from AIDA cohort	Kock et al.[34]	https://explore.data.humancellatlas.org/projects/35d5b057-3daf-4ccd-8112-196194598893
FinnGen	Kurki et al.[61]	https://www.finngen.fi/en/access_results
ENCODE	The ENCODE Project	https://www.encodeproject.org/
Human Cell Atlas scRNAseq data	Human Cell Atlas	https://cellxgene.cziscience.com/gene-expression
Super-Enhancer database 2.0	Wang et al.[42]	http://www.licpathway.net/sedbv2/
Software and algorithms		
Source code of noise calculation and enQTL analysis	This paper.	https://github.com/yxlong-science/enQTL-of-PBMC https://doi.org/10.5281/zenodo.17239294
Michigan Imputation Server	Das et al.[78]	https://imputationserver.sph.umich.edu/#!
PLINK v2.0	Purcell et al.[79]	https://www.cog-genomics.org/plink/2.0
coloc v5.2.3	Wang et al.[80]	https://github.com/chr1swallace/coloc
motifBreakR v2.23.0	Coetzee et al.[81]	https://github.com/Simon-Coetzee/motifBreakR
MASH	Urbut et al.[39]	https://stephenslab.github.io/mashr/
PEER	Stegle et al.[28]	https://github.com/PMBio/peer
tensorQTL v2.0.0	Taylor-Weiner et al.[27]	https://github.com/broadinstitute/tensorqtl
ANNOVAR	Wang et al.[40]	https://annovar.openbioinformatics.org/en/latest/
SCTransform v0.4.1	Hafemeister et al.[82]	https://github.com/satijalab/sctransform
SAMtools v1.17	Handsaker et al.[83]	https://github.com/samtools/samtools

METHOD DETAILS

Quality control of the genotype data

The genotype data were downloaded from Gene Expression Omnibus (GSE196830).[16] Intensity data files were transferred to PLINK data format using GenomeStudio PLINK Input Report Plug-in (v2.1.4).[79] We excluded the SNPs, if they 1) had missing rate >3%, 2) with minor allele frequency (MAF) < 0.01, 3) deviated from Hardy-Weinberg equilibrium ($p < 1 \times 10^{-3}$), 4) with ambiguous strand and flipped SNPs with reverse strand genotypes. We excluded the individuals, if they 1) had missing genotype rate >3%, 2) with abnormal autosomal heterozygosity (±3 SD deviated from the mean), 3) one of any pair of individuals with estimated relatedness larger than 0.125, 4) with non-European ancestry (defined as being ±6 SD from the European mean on PC1 and PC2 based on 1000G EUR individuals). The PCs were calculated using SMARTPCA software (v13050). Ultimately, we obtained 495,631 high-quality SNPs on 22 autosomes from 1,034 individuals.

Imputation of genotyping data

To enable the required formatting of imputation, we generated the bed and frequency file, executed QC script by HRC-1000G-check-bim.pl utility, and generated vcf file using VcfCooker (v1.17.5). The Michigan Imputation Server[78] was used for imputation with the Haplotype Reference Consortium panel (HRC r1.1 2016, EUR population) on each autosomal chromosome, with Minimac4 and Eagle v2.4.1. Finally, we obtained 5,295,853 SNPs with imputation quality $R^2 > 0.8$ and MAF >0.05 for subsequent analysis.

Expression data processing

We downloaded the processed data of OneK1K cohort from the Human Cell Atlas (HCA) portal. SCTransform (v0.4.1)[82] was used to normalize the gene UMI count matrix, accounting for the confounding effects from sequencing depth, pooling batch, library size, mitochondrial mapping percentage, and other latent factors. Since the calculation accuracy of gene expression noise is highly

correlated with the cell number of each cell type, we combined the published cell clustering and annotation into 8 cell types (CD4, CD8, NK, B, Mono, gdT, DC, and plasma).[16]

QUANTIFICATION AND STATISTICAL ANALYSIS

Gene expression noise calculation

To deconvolute the gene expression noise from the mean-variability dependence in single-cell data, we applied a quantitative statistical approach[3] with the following steps.

For the first step, we filtered the individuals if the cell counts of an individual in a specific cell type is less than 3 or the total number of expressed genes is less than 100. For all the included individual of a given cell type, we calculated the average expression and the observed expression square of coefficient of variation (CV^2) of each gene.

For the second step, we applied the gamma distribution of generalize linear model to fit the mean gene expression and CV^2 of all genes at individual level for each cell type. We defined a minimum value of gene expression (the 85^{th} percentile among all genes with $CV^2 > 0.3$) to exclude low-expressed genes with extremely high CV^2. We removed genes with abnormal CV^2 (±3 SD from the mean) as well. The starting intercept and slope were set as 0 and 0.5, respectively. Gamma fitting is the default parameter except using the identity link. To evaluate the fitting performance, we calculated the adjusted R^2 based on the sum of squares total, the sum of squares error, and the number of genes used for fitting.

The fitting curve of the gamma distribution represented the expected CV^2 of each gene and we calculated the residual-based noise by $\log_2 \frac{observed\ CV^2}{expected\ CV^2}$. If observed CV^2 is N/A (possibly caused by zero expression level in all cells), the residual is assigned as a missing value (N/A).

HVG and individual-level noise calculation

We estimated the statistical significance of deviation of each gene relative to expectation by the chi-square test and the resultant *p* values were adjusted by the Bonferroni correction (*FDR*). The HVGs were defined as the genes that were significantly deviated ($FDR \leq 0.05$, higher than expected) in more than 10% of individuals within a cell type. HVGs were calculated separately for each cell type. The individual-level noise was calculated by aggregating the noises of all genes within an individual.

enQTL identification

We used tensorQTL (v2.0.0)[27] for enQTL identification with covariates of the age and sex, the top six genotype PCs, and latent PEER factors (details in the following section). We restricted our search to SNPs within +/− 1Mb of transcription start site (TSS) of a gene. The nominal *p* value of lead SNP was extrapolated based on Pearson product-moment correlation and an adaptive permutation scheme by tensorQTL (parameter –permute 1,000, –maf_threshold 0.05). We corrected the nominal *p* value with the Benjamini–Hochberg procedure (method = "fdr"). Significance of the resulting SNP-enGene pairs was determined at a false discovery rate (*FDR*) threshold of 0.001 within each chromosome. To explore the conditionally independent signals of enQTLs within each locus, we performed the conditional analysis by regressing out the effect of the lead SNP. Specifically, we obtained the lead SNP of first round enQTL analysis and repeated the QTL mapping to identify the secondary enQTLs by adding the first-round lead SNP as a covariate. Conditional analysis was repeated up to five rounds. We focused on genes that were expressed in at least 10% of the individuals in each cell type. For eQTL, we applied the same procedure to enable fair comparison. Multivariate adaptive shrinkage (MASH)[39] was employed to assess enQTL sharing across cell types. An enQTL was classified as shared between two cell types if the local false sign rate (lsfr) of slopes was $\leq$0.1 in both cell types.

Latent factor identification and removal

We employed PEER (v1.0.0)[28] to identify the latent covariates of unwanted variance stemming from non-genetic factors. Prior to peer factor inference, the expression noise matrices were processed through inverse rank normalization. The K-Nearest Neighbors algorithm was applied to impute the missing values in the matrix of gene expression noise. Distance estimations were used to identify K individuals that are spatially similar or close to each other in the dataset. These K individuals were then used to estimate the missing value and to impute as the average of the K neighborhoods found in the dataset. We determined the number of PEER factors to include based on the number of identified enQTL and correlations among PEER factors (Figure S4), which were the top three, four, three, six, four, one, twenty PEER factors for CD4, CD8, NK, B, Mono, gdT and DC, respectively.

enQTL validation

To assess allele-specific expression noise in heterozygous individuals, we implemented a rigorous analytical workflow, as shown below. To generate BAM files and cell barcodes, raw FASTQ files were aligned using Cell Ranger (v9.0.1). To correspond to the cell annotations of OneK1K cohort, based on the cell annotations of the AIDA dataset in the original research, we got the cell annotations of these 149 individuals and merged the 28 cell annotations into 8 major immune cell subsets (CD4, CD8, NK, B, Mono, gdT, DC and Plasma). Given the technical limitation of 5′ scRNA-seq capturing, 36 enQTLs were retained for locating in the exonic or UTR regions. After applying filters to determine whether heterozygous individuals exist and whether expression is non-all-zero across

individuals, 16 testable enQTLs were retained for further analysis. Using mpileup function of SAMtools (v1.17),[83] we quantified allele-specific read counts for these enQTLs in heterozygous individuals. Using the read counts of two alleles, per-allele noise (CV^2) was calculated, effect directions were judged, and *pP*-value was estimated by two-sided Wilcoxon signed-rank test.

Cell-type-specific correlations between TF and enGene

To identify transcription factors (TFs) mediating enQTL cell specificity, we performed motif disruption analysis using lead enQTL SNPs from Table S10 via motifBreakR (v2.23.0).[81] This yielded 1,083 unique enQTL-TF-enGene regulatory trios. For each trio, we assessed cell-type-specific TF-enGene noise relationships using hierarchical linear mixed models (LMMs). Following model was fitted per TF-enGene pair: enGene noise$\sim$TF expression+celltype+(*S*|individual), and *S* represents the structure of random effects. All models were fit by maximum likelihood estimation. Significance of cell-type interactions was evaluated through likelihood ratio tests (LRTs) comparing full versus reduced models. Resulting *pP*-values underwent Benjamini-Hochberg FDR correction across all pairs, with $P_{adj} < 0.05$ defining significant interactions. For significant pairs, cell-type-specific slopes were computed via emmeans (v1.11.2) with FDR correction applied separately to interaction *pP*-values and slope *pP*-values within each pair.

Functional annotation and enrichment analysis of enQTL

The annotation of enQTLs and eQTLs involves two sections. First, we used the annotate_variation.pl script from ANNOVAR[40] to annotate all SNPs after imputation and filtering described above, applying the default annotation databases, except for -geneanno and -buildver hg19. Second, we collected publicly available epigenomic data for each immune cell line, mainly from two databases: the ENCODE Portal[41] and the Super-Enhancer database (SEdb).[42] From the ENCODE project, we downloaded pseudo-replicated peaks files (in narrowPeak format) of ATAC-seq and ChIP-seq (CTCF, H3K4me3, H3K9me3, H3K27me3, H3K36me3, H3K4me1, H3K27ac) and TAD and Loop files (in bedpe format) of intact Hi-C. From SEdb, we downloaded super-enhancer and enhancer files (in bed format). It is worth emphasizing that the data from ENCODE were based on the hg38 reference genome, so we used liftOver to convert them to the hg19 reference genome. The TAD boundary region is defined as the region extending 10 kb (two resolutions) on either side of the TAD boundary. Loops with an fdrBL $\leq$0.01 are considered significant loops, and ChIP-seq peaks with a Q-value less than 0.01 are defined as significant peaks.

The enrichment analysis of enQTL was performed in the following steps. First, we annotated lead SNPs of enQTL, respectively in different functional annotation categories (e.g., H3K4me3 peaks) based on their physical positions and quantified the proportion of enQTL SNPs in each category. Second, to ameliorate ascertainment bias, we randomly sampled at the same number of lead SNPs of enQTL (a single SNP per gene within +/− 1Mb of TSS) and quantified the proportion of enQTL SNPs in each category as the negative control. This sampling procedure was repeated 1,000 times. Third, we applied the Fisher's exact test in each functional annotation category as the ratio of the proportion of enQTL SNPs in a functional category over the mean proportion of the control SNPs in the category across 1,000 replicates, resulting in final *pP*-values and OR. The significant enrichment was defined as OR > 1 and $p \leq 0.05$. The same Fisher's exact test was used for the enrichment comparison between enQTLs and eQTLs.

Colocalization analysis between enQTL and GWAS loci

Colocalization analysis between enQTLs and GWAS signals was carried out using a Bayesian framework for colocalization (coloc).[80] A total of 29 hematopoietic traits and 5 autoimmune diseases were involved (full names and abbreviations in Table S12). In addition, 1,932 diseases from the FinnGen cohort were also included (https://www.finngen.fi/en).[61] First, we selected the GWAS loci at genome-wide significance of $p < 5 \times 10^{-8}$ and clumped by LD $R^2 < 0.1$ and region size of 100kb. Second, we selected the significant enQTL at *FDR* < 0.001 and overlapped with GWAS loci. Third, we performed colocalization for all enQTL-GWAS-overlapped loci. The colocalization statistics PP.H4 (posterior probability of both traits being associated with the same causal SNP) was calculated within the region of 100kb (±50 kb) of the GWAS lead SNP. PP.H4 > 0.7 were defined as the GWAS-QTL colocalized events. The colocalization of eQTL and GWAS was performed using the same procedure. Given the technical challenges in accurately mapping within the highly polymorphic MHC region,[84] we reported the colocalization results exclusively on non-HLA genomic regions.