

# Recalibrating differential gene expression by genetic dosage variance prioritizes functionally relevant genes

Philipp Rentzsch,<sup>1</sup> Aaron Kollotzek,<sup>1</sup> Kaushik Ram Ganapathy,<sup>2</sup>  
Pejman Mohammadi,<sup>3,4,5</sup> and Tuuli Lappalainen<sup>1,6,7</sup>

<sup>1</sup>Science for Life Laboratory, Department of Gene Technology, KTH Royal Institute of Technology, 17165 Solna, Sweden;

<sup>2</sup>Department of Integrative Structural and Computational Biology, Scripps Research Institute, La Jolla, California 92037, USA; <sup>3</sup>Center for Immunity and Immunotherapies, Seattle Children's Research Institute, Seattle, Washington 98101, USA; <sup>4</sup>Department of Pediatrics, University of Washington School of Medicine, Seattle, Washington 98105, USA; <sup>5</sup>Department of Genome Science, University of Washington, Seattle, Washington 98195, USA; <sup>6</sup>New York Genome Center, New York, New York 10013, USA;

<sup>7</sup>Department of Systems Biology, Columbia University, New York, New York 10032, USA

Differential expression (DE) analysis is a widely used method for identifying genes that are functionally relevant for an observed phenotype or biological response. However, typical DE analysis includes selection of genes based on a threshold of fold change in expression under the implicit assumption that all genes are equally sensitive to dosage changes of their transcripts. This tends to favor highly variable genes over more constrained genes where even small changes in expression may be biologically relevant. To address this limitation, we have developed a method to recalibrate each gene's DE fold change based on genetic expression variance observed in the human population. The newly established metric ranks statistically differentially expressed genes, not by nominal change of expression, but by relative change in comparison to natural dosage variation for each gene. We apply our method to RNA sequencing data sets from in vitro stimulus response and neuropsychiatric disease experiments. Compared to the standard approach, our method adjusts the bias in discovery toward highly variable genes and enriches for pathways and biological processes related to metabolic and regulatory activity, indicating a prioritization of functionally relevant driver genes. Tissue-specific recalibration increases detection of known disease-relevant processes. Altogether, our method provides a novel view on DE and contributes toward bridging the existing gap between statistical and biological significance. We believe that this approach will simplify the identification of disease-causing molecular processes and enhance the discovery of therapeutic targets.

[Supplemental material is available for this article.]

Since the advent of cDNA microarrays (Schena et al. 1995), differential gene expression profiling has been used to examine the characteristic gene regulatory changes of a specific phenotype, disease state, or perturbation. Statistical tests are used to determine whether the observed differences in expression of each single gene between groups of samples are statistically significantly differentially expressed (DE) over a chosen statistical threshold. Over the years, many different bioinformatics methods to test DE have been developed, adapting rigorous statistical tests to the particularities and assumptions of DNA microarrays and later RNA-seq. Some of the currently most popular methods for differential gene expression analysis are DESeq2 (Love et al. 2014), edgeR (Robinson et al. 2010), and limma-voom (Ritchie et al. 2015). These tools are designed to work with the count-based nature of RNA-seq data and incorporate normalization procedures to account for differences in sequencing depth and RNA composition across samples. Like most statistical tests, DE methods have been shown to sometimes produce false positives, particularly when analyzing RNA-seq data sets of large sample sizes. To address this issue, several different approaches of multiple testing correction have been developed (Benjamini and Hochberg 1995; Ge

et al. 2009; Ignatiadis et al. 2016) and are applied in many of the popular methods.

Whereas DE testing quantifies the statistical significance of DE, it is agnostic to its *biological relevance*, that is, whether the detected change in gene expression meaningfully reflects or contributes to changes in cellular functions. This means that a test's *P*-value does not carry any inherent meaning (Greenland et al. 2016; Wasserstein and Lazar 2016), and well-powered DE studies can result in hundreds, if not thousands, of DE genes. As a consequence, the actual expression fold change—the ratio of the expression levels between the two sample groups—is often considered an important secondary parameter (Cui and Churchill 2003; McCarthy and Smyth 2009; Zhang and Cao 2009; Jung et al. 2011; Harrison et al. 2019). Many DE studies apply an arbitrarily chosen minimum fold change threshold and inspect only statistically significant genes surpassing this cutoff. This makes sense insofar as the magnitude of an expression change is obviously relevant (Naqvi et al. 2023) and, for a single gene, bigger expression changes tend to matter more than smaller ones. However, the absolute fold change in itself is of limited relevance when comparing changes between different genes because different genes have different levels of dosage constraint (Rice and McLysaght 2017). This means that a measured change in expression in one

**Corresponding authors:** philipp.rentzsch@scilifelab.se,  
tuuli.lappalainen@scilifelab.se

Article published online before print. Article, supplemental material, and publication date are at <https://www.genome.org/cgi/doi/10.1101/gr.280360.124>. Freely available online through the *Genome Research* Open Access option.

© 2025 Rentzsch et al. This article, published in *Genome Research*, is available under a Creative Commons License (Attribution 4.0 International), as described at <http://creativecommons.org/licenses/by/4.0/>.

gene may be comparable to biological noise and may fall within the spectrum of natural variation in the population. However, a deviation of the same magnitude in another gene can be highly unusual and lead to immediate cellular consequences, either by affecting other genes or by directly altering the phenotype. Such a change could hence be described as not only statistically but also biologically significant. Prior work shows that genes with constrained expression are enriched for drivers of cellular processes and disease (Lek et al. 2016; Rice and McLysaght 2017; Mohammadi et al. 2019; Karczewski et al. 2020; Collins et al. 2022; Dong et al. 2023). It follows that the subsequent difference in gene responsiveness makes nominal fold change a poor proxy for biological relevance, an effect that may contribute to the systematic differences between gene expression changes and associations between genetic variants and traits (Mostafavi et al. 2023).

The best studied quantity that describes how tolerant genes are to dosage variation is haploinsufficiency, the gene's intolerance to heterozygous deletion or loss-of-function variants. A related concept of triplosensitivity refers to intolerance to duplication. Beyond these large changes that affect one entire copy of a gene, especially noncoding genetic variants affect the expression of nearby genes. For many genes, as recently exemplified by the gene encoding the transcription factor SOX9 (Naqvi et al. 2023), these regulatory changes are associated with differences in phenotype and disease risk (Albert and Kruglyak 2015). To our knowledge, there exist currently no assays that measure the viable dosage range of mRNA expression for each gene in a genome. However, it has been shown that haploinsufficiency of genes aligns well with the amount of purifying selection in the genome (Lek et al. 2016; Karczewski et al. 2020; Collins et al. 2022). Indeed, depletion of genetic variation in populations has been a powerful genome-wide approach for detecting molecular variation that is not tolerated by natural selection.

In this study, we have developed a novel calibration method to measure DE by scaling standard expression changes with the population genetic dosage variation estimate  $V^G$ , a metric which we had previously introduced (Mohammadi et al. 2019). This calibration approach was tested and validated in multiple previously published data sets, and the resulting recalibrated gene expression fold changes were evaluated for their ability to improve the interpretability of DE analyses across diverse data sets and tissues. The objective of this method is to provide a more accurate framework for identifying genes that play a pivotal role in regulating cellular processes and disease mechanisms.

## Results

### Variance in gene expression as a fold change recalibration metric

The amount of population variability in expression is different for each gene. Whereas some of the measured variability is due to technical factors, physiological differences, or external influences from the environment, some of it is due to genetic differences between individuals. To infer the component of natural variation that is due to genetic *cis*-regulatory differences, we previously introduced ANEVA, a method for assessing the gene expression distribution based on allelic expression (AE) of heterozygous SNPs or eQTLs (Mohammadi et al. 2019). Recently, we have further extended this principle to use allelic expression over entire haplotypes (ANEVA-h), improving robustness and data coverage (K Ganapathy, M Broly, S Silverstein, et al., in prep.; <https://zenodo.org/records/15226575>).

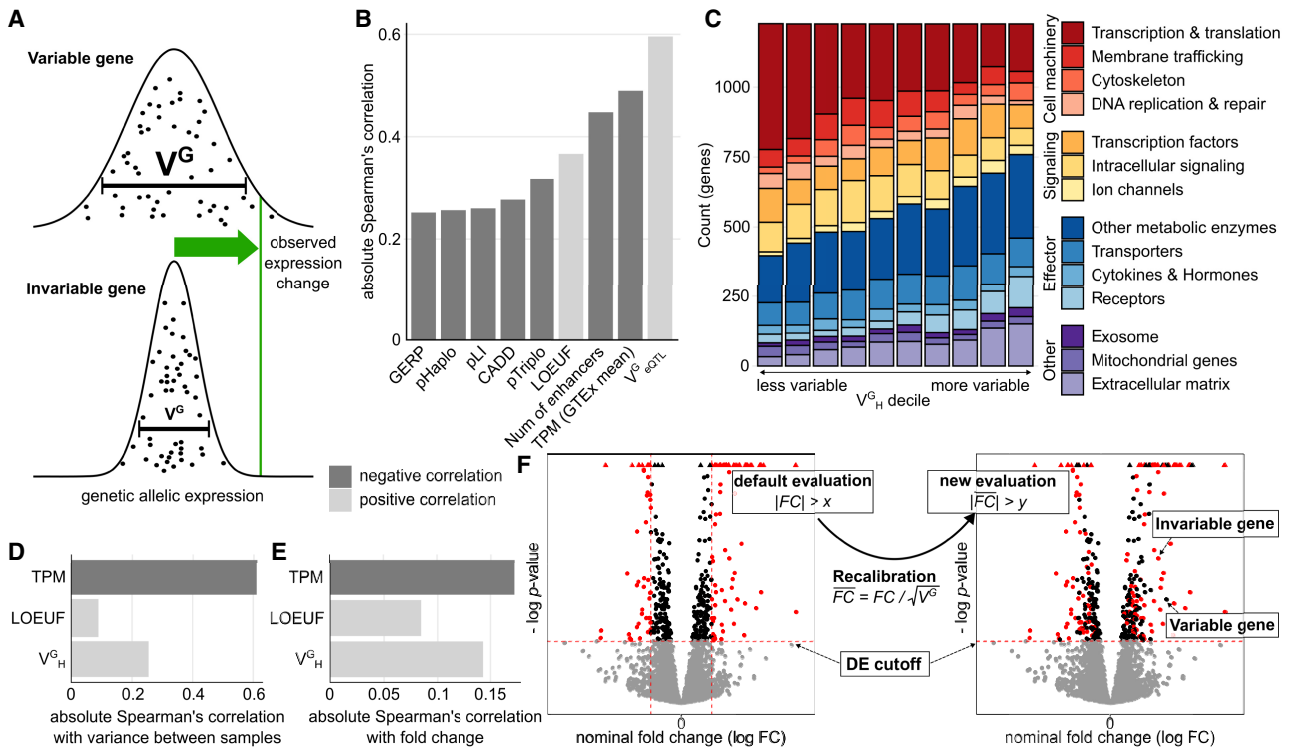
Applying ANEVA-h to GTEx v8, an RNA-seq data set from more than 800 individuals, we derived an estimate of gene expression variance ( $V^G$ ) for nearly all human genes. We refer to these estimates as  $V^G_H$ . Because both expression and variability depend on body tissue and cell type,  $V^G_H$  is initially calculated separately for 50 different tissues in GTEx. An overall  $V^G_H$  value is then calculated as the weighted harmonic mean across tissues.

As  $V^G$  is based on the allelic fold change (Mohammadi et al. 2017) of genetic regulatory effects, its unit is analogous to log fold changes used in DE analyses. The magnitude of  $V^G$  describes the range in which observed gene expression changes fall within the population (Fig. 1A). Large values imply greater variability and thus a wider range of expression levels in the population.  $V^G$  hence serves as a parameter to classify an expression change relative to the respective gene's mean, that is, where a sample with the observed change would fall within the natural (population) range.

In our previous publication (Mohammadi et al. 2019), we showed that AE-based  $V^G$  ( $V^G_{AE}$ ) of a gene is correlated with selective constraint. Known haploinsufficient genes have lower  $V^G_{AE}$  on average, and  $V^G_{AE}$  is correlated with other constraint metrics such as conservation scores or the probability of being loss-of-function intolerant (pLI). Here, we measured the correlation of our most recent, haplotype derived  $V^G_H$  estimates to additional gene constraint metrics (Fig. 1B). One of the most highly correlated metrics ( $\rho_{\text{Spearman's}} = 0.36$ ) is the loss of function tolerance metric LOUEF (Karczewski et al. 2020). Most interesting are the haploinsufficiency and triplosensitivity metrics pHaplo and pTriplo (Collins et al. 2022). Whereas  $V^G_H$  is less correlated with both metrics than, for example, LOEUF, we found that, in contrast to the latter,  $V^G_H$  is slightly more correlated with pTriplo ( $\rho_{\text{Spearman's}} = -0.31$ ) than pHaplo ( $\rho_{\text{Spearman's}} = -0.25$ ), indicating its potential in capturing sensitivity to both up- and downregulation. The high negative correlation with the number of enhancers ( $\rho_{\text{Spearman's}} = -0.46$ ) suggests that the genetic population variance in expression is linked to buffering at the molecular level (Wang and Goldstein 2020).

In addition to metrics of constraint, we found that  $V^G$  estimates are associated with gene function. As described for a simpler GTEx-derived "allelic Fold Change" metric (Dong et al. 2023), genes annotated as part of the central cellular processes, particularly those involved in transcription and translation, are enriched among genes with low  $V^G_H$ . In contrast, effector proteins such as receptors, as well as extracellular matrix proteins, are more likely to have a high  $V^G_H$  (Fig. 1C).

We next investigated how  $V^G$  is correlated to gene expression variance in experimental data. To this end, we analyzed a data set of 81 human macrophage samples that were treated with IFNG (Alasoo et al. 2018). As expected due to the well-documented mean-variance relationship, expression level (transcripts per million [TPM]) was highly correlated with interindividual variance between control samples, but also  $V^G_H$  and LOEUF were correlated to variance (Fig. 1D), which is not unexpected for different metrics of population variance. More notably, the expression fold change upon IFNG stimulus was also correlated with  $V^G_H$  ( $\rho_{\text{Spearman's}} = 0.143$ , 95%-confidence interval  $[CI_{95}] = [0.125; 0.161]$ ) (Fig. 1E), significantly more so than LOEUF ( $\rho_{\text{Spearman's}} = 0.085$ ,  $CI_{95} = [0.066; 0.103]$ ) and within the range of TPM ( $\rho_{\text{Spearman's}} = -0.172$ ,  $CI_{95} = [-0.190; -0.155]$ ). Whereas DE significance testing accounts for the variance in the studied data set, the correlation between fold change and  $V^G_H$  indicates that genes that are more likely to



**Figure 1.** Fold change recalibration using  $V^G$ . (A) Depending on variability of a gene, an observed fold change may be within the natural variation that is observed in the general population.  $V^G$  is an estimate of this population variance. (B) Correlation of  $V^G_H$  with other gene metrics. Negative correlations are colored dark gray; positive correlations are light gray. (C) Distribution of functional categories in genes split by  $V^G_H$  decile. Some functional categories like transcription- and translation-associated genes were enriched for small  $V^G_H$  estimates, and others, like receptors, were enriched for high  $V^G_H$  estimates. (D) Correlation of gene metrics with experimental gene expression variance. Data sourced from Alasoo et al. (2018) studying the effects of IFNG treatment on 81 human macrophage samples. (E) Correlation of gene metrics with absolute gene expression changes in the IFNG experiment. (F) Concept of gene expression recalibration using  $V^G$ . As a result of this process, genes are prioritized not by nominal log fold change ( $FC$ ) but by recalibrated log fold change ( $\overline{FC}$ ).

be noticeable outliers in a DE study also have a higher population variance in general.

The idea behind the original ANEVA method was to have a test that evaluates whether the genetic effect in a gene in one particular sample is an outlier compared to the genetic variance in the population. Conversely, here, we propose to use the derived variance metric  $V^G$  to rescale nongenetic gene expression differences between two groups in order to compare expression changes across different genes on the same scale, relative to the population variance in each. Whereas the above findings provide support for using  $V^G$  to rescale gene expression fold changes, the main motivation for  $V^G$  as a metric itself is unit equivalence. Unlike scores such as LOEUF, GERP, CADD, or pLI, where there is no direct relationship between score value and variability, a gene with a  $V^G$  of 0.02 is twice as variable as a gene with a  $V^G$  of 0.01. We used this property to recalibrate the experimental log fold change by scaling it by the average change in expression. This average change in expression is the standard deviation, which is the square root of the variance. Thus, with this approach, which we call recalibration, we standardize the observed fold change for each gene with the standard deviation of the genetically regulated gene expression given by the square root of  $V^G$  (Fig. 1F). To distinguish between pre- and postrecalibration, we refer to log fold changes obtained from an experiment as “nominal” and fold changes after rescaling as “recalibrated.” Recalibration is an additional analysis step after significance testing, and it changes the relative order of genes assigned as significant based on standard DE testing. This means

that small changes in expression for some genes may be considered relatively more significant than the same change would be for other genes.

### Recalibration shifts the analysis focus

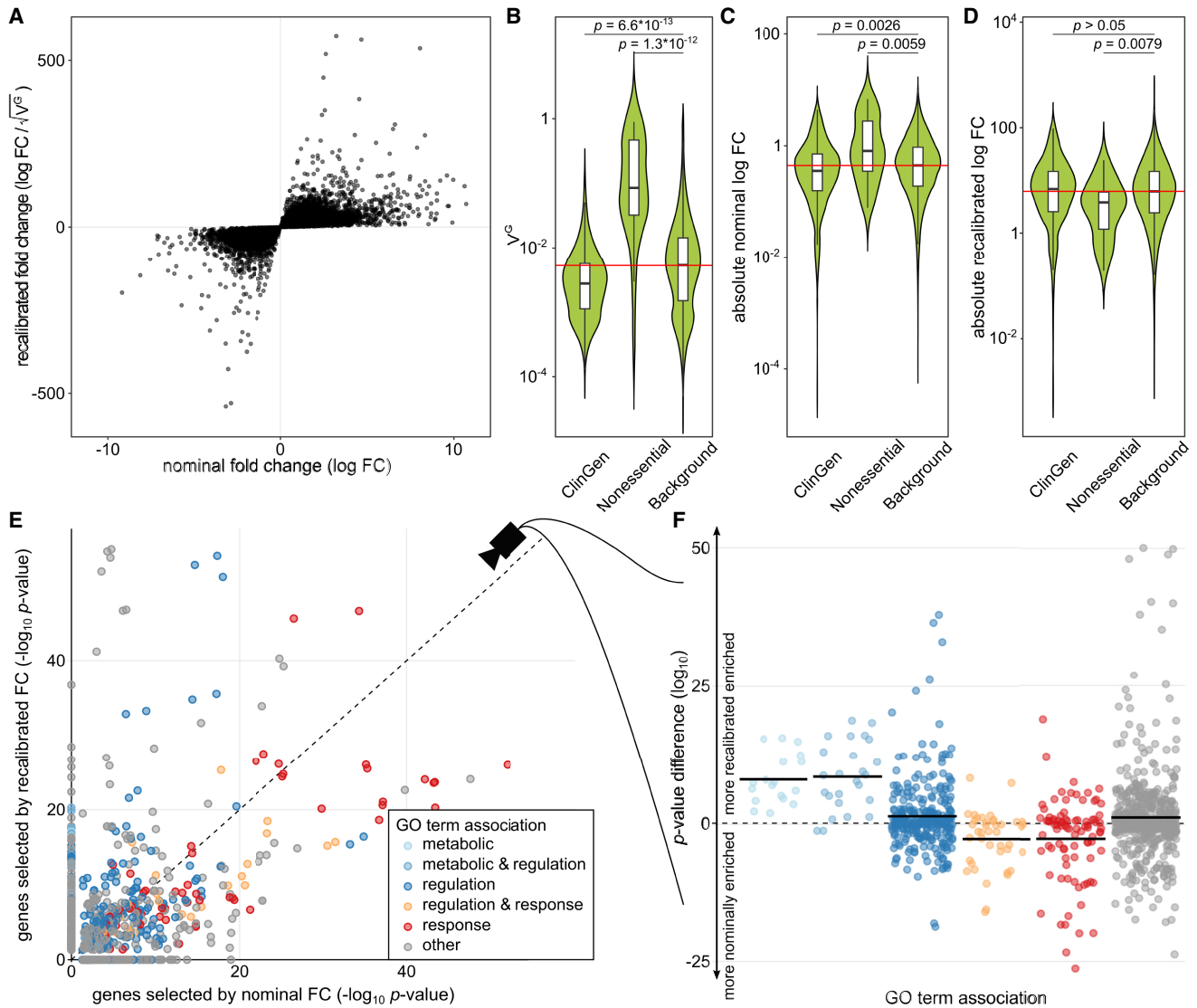
To assess the effect of a recalibration with  $V^G$ , we used the previously introduced data set of 81 human macrophage samples stimulated with IFNG. After processing the samples and selecting DE genes with a strict false discovery rate of less than 0.001, the data set yields 20,402 DE genes (out of 29,438 in the data set). Of these, 5573 genes also have an absolute FC greater than 1. The data set hence serves as an excellent illustration of the need for selection criteria beyond statistical significance to identify the most relevant genes for further analysis. To compute recalibrated fold changes, we restricted further analysis to the 22,650 genes in the data set for which  $V^G_H$  estimates have been calculated. Of these, 17,494 were DE and 4586 DE genes had an aFC greater than 1. The recalibrated fold changes were significantly correlated with the nominal fold changes (Pearson’s correlation of absolute values: 0.52) (Fig. 2A).

To examine the effect of the recalibration, we first turned to sets of well-characterized genes. Known haploinsufficient genes from ClinGen on median had lower than average  $V^G_H$  estimates, whereas those of nonessential genes were generally higher than average (Fig. 2B). Nonessential genes had higher absolute fold changes upon IFNG stimulus than ClinGen and background

genes, indicating that a standard DE analysis might easily focus on these functionally irrelevant genes (Fig. 2C). The recalibration results in an adjustment of the fold change ranges for the different gene sets (Fig. 2D), with their respective distributions no longer being significantly different.

Using the 2000 most DE genes by nominal and recalibrated absolute FC, respectively, we observed a shift in GO term enrichments (Fig. 2E; Supplemental Table S1). In particular, terms matching the string “response,” like “immune response” (GO:0006955) and “response to other organism” (GO:0051707), were less en-

riched, whereas terms matching “regulation” like “regulation of biological process” (GO:0050789), or “metabolic” like “metabolic process” (GO:0008152), were more enriched among the genes selected after recalibration (Fig. 2F). This effect was robust to the number of genes analyzed (Supplemental Fig. S1) or when the gene set enrichment was calculated based on gene ranks (GSEA) (Supplemental Fig. S2). However, we note that many GO terms represent broad categories of genes rather than specific pathways, and GO term enrichments are not independent of each other as most genes are associated with many different GO terms. We



**Figure 2.** Fold change recalibration of a DE experiment of IFNG treatment. (A) Nominal compared to recalibrated fold changes of individual genes from Alasoo et al. (2018). (B–D) Impact of recalibration on genes from three different gene sets: ClinGen, Nonessential, and all other genes (Background). Part B displays the distribution of  $V^c_H$  per set; part C, the absolute log fold changes before recalibration (nominal); and part D, the recalibrated absolute log fold changes. Each violin represents the total distribution of values, with boxes in the violin indicating 25% and 75% quantiles and the median illustrated as a black line. Whiskers are 1.5 times the interquartile range. The red line denotes the median over all genes. Significance labels are the result of a two-sided comparison with Mann–Whitney  $U$  test of the two groups. (E) Effect on GO term enrichment: adjusted  $P$ -values of GO term enrichment from 2000 genes selected by nominal compared to recalibrated FC. GO-terms are colored based on string matching (association). The dashed center line marks the same enrichment in both sets. GO terms above the line are more enriched after the recalibration, and terms below the line are more enriched before. (F) Different view on the change in GO term enrichment  $P$ -values grouped by associations. Colored dots are GO terms; the colored vertical lines are the log-mean per association. Terms containing the words “regulation” (blue) and “metabolic” (light blue) are, on average, more enriched after recalibration, whereas terms containing the word “response” (red) are, on average, more enriched before recalibration.

analyzed the extent of sharing by clustering the GO terms by the genes associated with each term, finding that the two main clusters including GO terms whose names contain the word “response” are, on average, less enriched after recalibration, whereas 10 out of 12 other GO term clusters where multiple terms match the word “regulation” are more enriched after recalibration (Supplemental Fig. S3). However, it should be noted that, whereas the differences in enrichment can be large and only 991 genes are in the top 2000 in both lists, the change in the number of genes per GO term is small (Supplemental Fig. S4). We found the same enrichment of GO terms matching “regulation” and “metabolic” when performing the same analysis on experiments of macrophages perturbed with *Salmonella* and *Salmonella*+IFNG from the same experimental data source (Supplemental Fig. S5). We also found the same trend when recalibrating using eQTL-based  $V^G$  estimates (Supplemental Fig. S6). All these results indicate that nominal fold change captures *response* to stimuli, whereas recalibration enriches for *regulators* of cellular response.

Next, we wanted to analyze whether this pattern holds for a DE study of a more limited size than the well-powered Alasoo et al. (2018) data. To this end, we analyzed the data from Findley et al. (2021) with multiple in vitro stimuli of three different cell lines in six biological replicates, focusing on the response to metal ions. A known key mechanism in this response is a family of proteins, metallothioneins, that have the ability to bind metal ions, providing protection against metal toxicity (Sekovanić et al. 2020). In the stimulus experiments of copper and zinc ion solutions, the top 10 DE genes based on classical fold change include five and seven of the total of nine metallothionein genes, respectively (Supplemental Figs. S7A, S8A), and the associated GO terms “cellular response of copper/zinc ion” (GO:0071280) are significantly GO enriched. Recalibration reduces the metallothionein gene ranking and GO enrichment (Supplemental Figs. S7B, S8B), as the  $V^G_H$  values of the metallothionein genes are in the top quartile of all genes.

Altogether, these analyses show how the choice of DE approach depends on the goals of the analysis: the standard approach readily picks up genes that respond to the given stimulus. However, recalibration provides a gradual adjustment to DE gene rankings, which may help to deprioritize highly variable responder genes to highlight molecular regulators and drivers of cellular response.

### Tissue-specific expression fold change recalibration

The analyses presented so far use mean  $V^G_H$  values across tissues, which does not account for tissue-specific differences in variability. The logical alternative is to use tissue-specific  $V^G$  estimates of the closest tissue available. However, whereas the mean  $V^G_H$  estimates cover 26,760 genes, only between 992 and 21,051 genes had estimates in each of the 50 tissues (Fig. 3A). Based on the previous finding that estimates of  $V^G$  between tissues correlate with tissue expression (Mohammadi et al. 2019), we developed an inference method of estimating  $V^G$  based on the expression in other, correlated tissues (Supplemental Fig. S9; Methods). We refer to the generated  $V^G$  estimates as inferred ( $V^G_I$ ).  $V^G_I$  is, for most tissues, more correlated to  $V^G_H$  (median 0.82) (Fig. 3B) than the most strongly correlated other GTEx tissue (median 0.77), with the exception being tissues like the two GTEx skin tissues that are much more similar to each other than any other tissue. Due to the simplicity of the method, it allows us to generate  $V^G_I$  estimates for all 26,760 genes in all 50 GTEx tissues. To recalibrate

as many genes as accurately as possible, we combined  $V^G_{H, \text{TISSUE}}$  and  $V^G_{I, \text{TISSUE}}$  into  $V^G_{HI, \text{TISSUE}}$ , with  $V^G_H$  taking precedence.

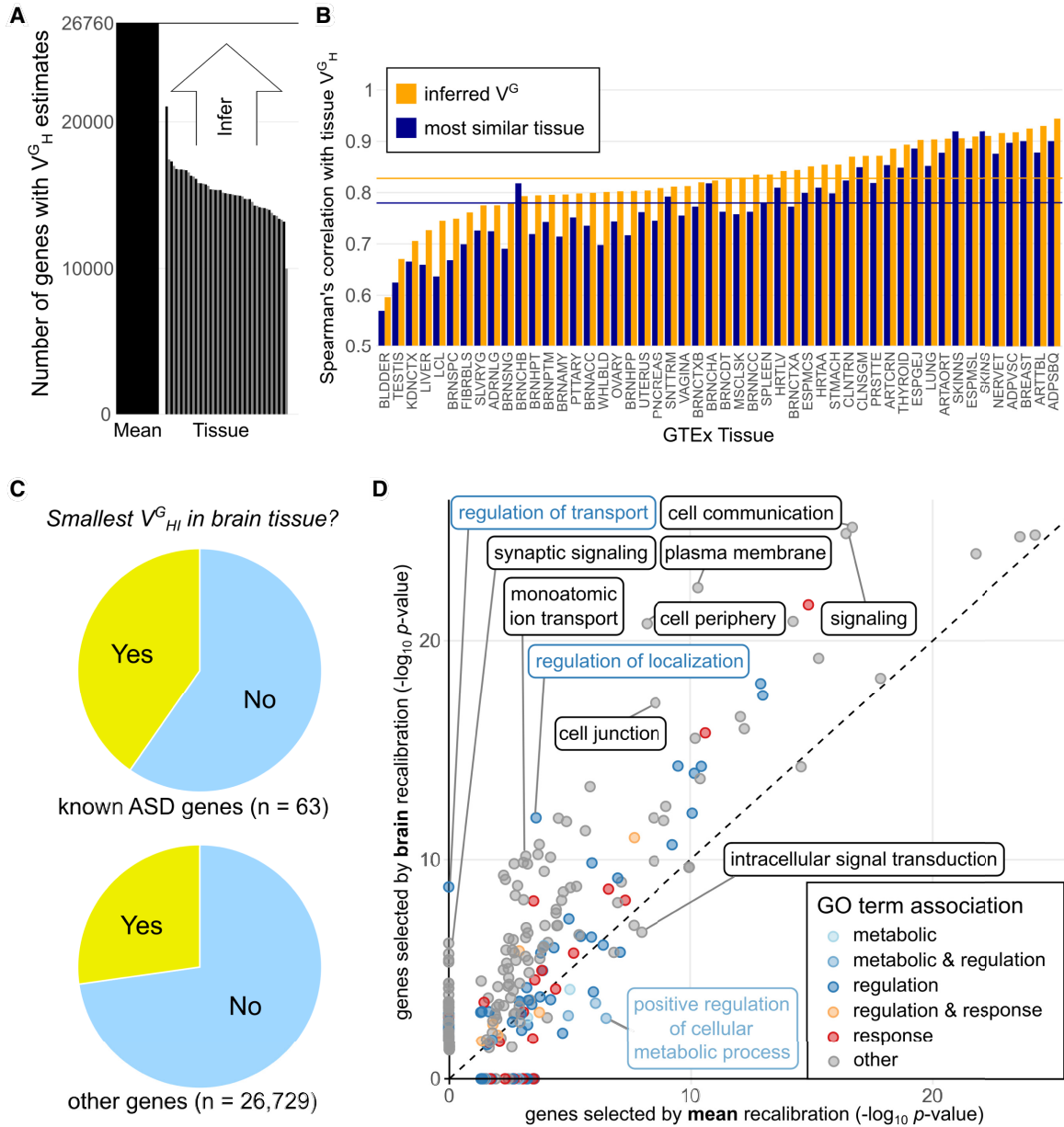
Tissue-specific  $V^G_{HI}$  estimates allow for experiment-specific recalibration in order to highlight genes under constraint in that tissue. For example, we find that a significantly increased fraction of previously known autism spectrum disorder (ASD) genes (An et al. 2018) has the lowest  $V^G_{HI}$  in one of the 14 GTEx brain tissues (binomial test  $P$ -value of  $5.2 \times 10^{-4}$ ) (Fig. 3C). This indicates that  $V^G$ , unlike genetics-based scores, can capture tissue-specific regulatory constraint. To explore this further, we used the psychENCODE data set (Gandal et al. 2018), which studied ASD, schizophrenia (SCZ), and bipolar disorder (BD) across hundreds of RNA-seq samples primarily from the prefrontal cortex.

The complete psychENCODE data set measured the expression of 25,772 genes. For recalibration, we used the  $V^G_{HI}$  estimates for 19,164 genes of the GTEx Brain - Frontal Cortex tissue. Of the 1611 statistically significant DE genes in ASD versus controls, 1444 have  $V^G$  estimates and can be recalibrated. First, we replicated the general pattern of recalibration enrichment for regulatory GO categories compared to nominal fold change ranking, analogous to those in the IFNG stimulus data (Supplemental Fig. S10; Supplemental Table S2). Importantly, GO term enrichments among the top 700 genes after recalibration with tissue-specific  $V^G$  estimates compared to mean  $V^G$  showed improved results (Fig. 3D), detecting enrichments for GO terms such as “regulation of transport” (GO:0051049) and “synaptic signaling” (GO:0099536), which were not significantly enriched when selecting genes based on recalibration with mean  $V^G$  estimates. Synaptic signaling is well-characterized as a key process in ASD (Jiang et al. 2022), whereas regulation of transport contains genes which modulate the transport of various molecules across cellular membranes. Using the tissue-specific  $V^G$  for recalibration, we observed similar increased enrichments of neural activity-linked GO terms like cell communication and signal transduction for both SCZ (Supplemental Fig. S11) and BD (Supplemental Fig. S12).

## Discussion

Here, we have introduced a novel approach for one of the most common analyses in computational biology: ranking of DE genes. Using a genetic regulatory variance metric for recalibration—as opposed to metrics relying on transcriptome or genetic data alone—has many advantages. Firstly, it provides a robust and generalizable estimate of biological constraint on expression that is not confounded by intermingled technical and biological noise in transcriptome studies. Most constraint scores, particularly those based on genetic variation, do not capture tissue specificity, which we have shown here to further increase the enrichment of likely regulatory pathways. Furthermore, using *genetic* variance to better interpret environmental or other *nongenetic* differences brings together two typically distinct areas of biological enquiry. Whereas a simpler method (Starr et al. 2023) has been proposed for assessing whether an observed expression change is different from the natural variability,  $V^G$  is based on allelic fold changes (Mohammadi et al. 2017) that are measured in the same unit as DE fold change. Thus, this is the first method to compare different genes from an experiment on the same, biologically interpretable scale.

We have shown that recalibration deprioritizes highly variable genes and removes the bias of DE results, easily highlighting nonessential genes with high population variation. Reranking of DE genes led to increased enrichment of molecular processes related to regulation and correctly identified key driver processes in



**Figure 3.** Tissue-specific recalibration. (A) Number of genes  $V_G^H$  is estimated for, both as the tissue agnostic mean and per tissue. (B) Per tissue correlation of tissue  $V_G^H$  estimates with the inferred  $V_G^I$  estimate, compared to correlation with the most highly correlated tissue. Horizontal lines are medians for all tissues. (C) Comparison in which tissue the  $V_G^H$  of a gene is lowest. Compared to all other genes, known ASD genes are significantly more likely to have the lowest  $V_G^H$  in one of the GTEx brain tissues. (D) Impact on GO term enrichment of using tissue-specific  $V_G^H$  estimates (GTEx tissue Brain Frontal Cortex) compared to mean  $V_G^H$  for recalibration in a DE experiment by PsychENCODE comparing ASD patients and controls. These results indicate that high-coverage tissue-specific metrics of regulatory constraint from the new haplotype-based  $V_G^I$  estimates in GTEx increase the detection of biologically meaningful expression signatures in disease.

neuropsychiatric traits, although the lack of broadly applicable gold-standard annotations of gene function beyond GO categories limits the resolution of these insights. Indeed, whereas we show that recalibration produces robust patterns for gene sets, users are encouraged to interpret results from individual genes with some caution. Nevertheless, our results indicate that our approach has the potential to identify phenotypic drivers, guiding downstream analyses toward genes and processes that are biologically causal. These would represent potential targets for interventions. Conversely, the standard DE approach is sensitive in picking up

genes that reflect the response to the signal of interest. Whereas such genes do not necessarily drive downstream processes, they may be informative—for example, as biomarkers of external stimulus. Thus, the standard and recalibrated DE analyses highlight different types of biology, both of which may be of interest. An additional interesting group of genes that our current approach is not particularly well suited to highlight include genes where high genetic and environmental variability contributes to adaptation to different environments (Kelly et al. 2012). However, the ability to focus the downstream analysis to phenotypic drivers

and deprioritize highly variable genes that merely reflect the response will be valuable in many applications.

The approach presented here has some limitations, many of these derived from the input data for  $V^G$  estimates. Despite our innovations in deriving  $V^G$  estimates for a substantially increased number of genes, especially in tissue-specific analyses, many genes still lack an empirically derived  $V^G$  estimate, although this can be successfully alleviated with the presented inference methods. Missing genes are generally lowly expressed, which is the reason why allele-specific data (and similarly eQTLs) could not be determined and that low-expressed genes are generally more variable and thus negatively affected by recalibration. Thus, missing genes are likely to be negatively correlated to functional importance, unlike for some genetics-based scores (Lek et al. 2016). Furthermore, allelic analysis is highly sensitive to potential clonality in cell lines, making some data sets unsuitable for  $V^G$  estimation.

Another key aspect is the context specificity of  $V^G$ . Our previous findings indicate that genetic background has only a minor effect on gene expression variability (Mohammadi et al. 2019), although it is unclear if there are more subtle differences and exceptions that cannot be captured with current data sets. Our results both highlight the importance of tissue-specific  $V^G$  and show a high correlation of closely related tissues, likely deriving from some strong genetic *cis*-regulatory effects being shared between tissues (The GTEx Consortium 2020). This is likely to extend to different cell states, where gene-environment (G×E) interactions—that is, response eQTLs—might create situations where  $V^G$  calculated on baseline data will not fully capture genetic variability under a different cell state. However, major G×E effects affect a relatively small number of genes (Manuck and McCaffery 2014; Kim-Hellmuth et al. 2017; Wang et al. 2019). Thus, recalibration can be done with tissue, cell type, or cell state proxies, but growing gene expression data sets will further improve the quantity and quality of  $V^G$  estimates that closely match the differential expression data, which is likely to improve the quality of the recalibration.

Our results demonstrate that the current  $V^G$  scores are of sufficient quality for our straightforward recalibration to yield biologically meaningful signals. This evidence will hopefully inspire future incorporation of  $V^G$  noise estimates in the fold change recalibration step and potentially even into DE significance testing itself to account for general population variance. In such analyses, accurate matching of the cell type and sometimes even the population background between the experiment and  $V^G$  values is likely to become an even higher priority.

Recalibration changes the unit of differential expression from molecular fold change to how many standard deviations an expression change is from the population mean. Neither approach has a natural threshold of biological significance, and we prefer the ranking approach where top N genes are inspected. The ultimate way to identify molecular changes that are biologically meaningful are experimentally intensive, but an exciting future direction, will be to map the complex and often nonlinear relationships of molecular expression changes, downstream phenotypic consequences, and population variance (Naqvi et al. 2023; Domingo et al. 2024).

In summary, we have introduced a new approach for prioritizing potential regulatory drivers in DE analyses. This provides yet another demonstration of the value of population-scale RNA-seq data for enhancing biological interpretability of broadly used study designs and enabling new discoveries in the future.

## Methods

### Genetic variance in gene expression

$V^G$  is a metric that estimates the genetic variance in expression for genes in the human population, which we introduced in previous publications (Mohammadi et al. 2017, 2019). There, we described approaches based on different data types to calculate  $V^G$ : from allelic expression based on single SNVs (AE), from allelic expression based on haplotype information (H), and from expression quantitative trait loci (eQTL). Briefly, AE-based  $V^G$  ( $V^G_{AE}$ ) is derived based on the comparison of allele read counts for each single nucleotide variant (SNV) in the two haplotype copies in a diploid individual. This information is aggregated by SNV frequency over all individuals in the tested population and fitted via ANEVA (analysis of expression variation), a model for population AE data, to infer with  $V^G_{AE}$  the level of genetic expression variability between haplotypes. Haplotype-based  $V^G$  ( $V^G_H$ ), in contrast, is derived by aggregating read counts across phased haplotype blocks. Here, we are using the haplotype-expression matrices from GTEx v8, that were generated with phASER using WASP corrected RNA-seq data (Castel et al. 2020). This aggregation increases the total read counts and enhances the number of genes with detectable allelic expression. The aggregated data are then fitted with a simplified version of the ANEVA model, replacing the aggregation of probabilities over multiple SNVs with a single binomial-logit-normal probability estimator to infer  $V^G_H$  (K Ganapathy, M Broly, S Silverstein, et al., in prep.; <https://zenodo.org/records/15226575>). In contrast,  $V^G_{eQTL}$  is generated from the most significant eQTL of a gene as an aggregate of effect size and allele frequency. All approaches of generating  $V^G$  have been applied to RNA-seq samples from the GTEx project, generating separate estimates for each GTEx tissue. A mean  $V^G$  has been calculated as a weighted harmonic mean over all tissue-specific  $V^G$  estimates, using expression per tissue, measured in TPM, as weight.

$V^G_H$  estimates used in this study were generated on data from GTEx v8 and were reused from K Ganapathy, M Broly, S Silverstein, et al., in prep. (<https://zenodo.org/records/15226575>).  $V^G_{eQTL}$  estimates are based on GTEx v7 and were reused from the Mohammadi et al. (2019) publication. Genes for which no  $V^G$  has been estimated were excluded from the analyses. All  $V^G_H$  estimates are provided in the supplement (Supplemental Table S3).

### Recalibrating expression fold change

In order to prioritize biological significance, expression fold change (FC) from DE analysis is standardized relative to the standard deviation of genetically regulated gene expression:

$$\overline{FC}_g = \frac{FC_g}{\sqrt{V_g}}, \quad (1)$$

where  $FC_g$  and  $\overline{FC}_g$  are the nominal and recalibrated log fold change of a gene  $g$ , and  $V_g$  is  $V^G$ , the expected variance in gene dosage introduced by genetic variation in the population. An implementation of the process can be found in Supplemental Code S1.

### Other gene metrics

RVIS scores per gene were obtained from Petrovski et al. (2013). ncCADD, ncGERP, and ncRVIS were obtained from Petrovski et al. (2015). The numbers of enhancers and super-enhancers per gene were derived from GeneHancer v5 (Fishilevich et al. 2017). Based on the GFF file with all enhancers from genecards.org, enhancers with a score  $\geq 0.7$  and a gene association “link\_score”  $\geq 7$  were filtered. The final numbers represent the total counts after filtering per gene. The index for tissue-specific expression in

GTEX tau (Palmer et al. 2021) was downloaded from [genomics.senescence.info/gene\\_expression/Tau\\_score.zip](https://genomics.senescence.info/gene_expression/Tau_score.zip). The triplosensitivity and haploinsufficiency scores, pTripto and pHaplo, respectively, were obtained from the supplemental data of Collins et al. (2022). The same data set was further used to obtain EpiScore (Han et al. 2018), pLI (Lek et al. 2016) and LOEUF (Karczewski et al. 2020) scores were obtained from gnomAD (<https://gnomad.broadinstitute.org>), using data from the v3 release. Transcripts per million values for each tissue were downloaded from GTEx (<https://gtexportal.org/home/>), using GTEx version 8. When referring to TPM in the text, we are generally using the mean TPM across all GTEx tissues, which were, in turn, calculated as the median across all individual samples of each tissue.

To account for the different scales of different gene metrics, correlations with  $V^G$  were calculated as Spearman's correlation, and 95% confidence intervals of correlation were determined using bootstrapping.

The list of nonessential genes was obtained from Zarrei et al. (2015). ClinGen genes were downloaded via the dosage sensitivity curated gene list from [ftp://clinicalgenome.org/ClinGen\\_gene\\_curation\\_list\\_GRCh38.tsv](ftp://clinicalgenome.org/ClinGen_gene_curation_list_GRCh38.tsv) (version from 2023/02/13). Haploinsufficient genes were selected by joining all genes with a haploinsufficiency score of 1, 2, or 3. The background gene set contains all genes with H-based  $V^G$  values, except for the genes included in one of the two other gene sets. Genes with a TPM of <1 were filtered.

### KEGG gene annotations

The annotation of KEGG functional categories was adapted from Dong et al. (2023). We reused their functional category labels but excluded the category "Domain-containing proteins not elsewhere classified" due to frequent overlaps with other categories. When multiple labels could be applied to a gene, we prioritized the one that appeared least frequently in the list of all genes with  $V^G$ . Genes that were not found in KEGG or that did not match any of the labels were omitted.

### RNA expression data processing

RNA expression data from Alasoo et al. (2018) were downloaded from Zenodo (<https://zenodo.org/record/839011> and <https://zenodo.org/record/4678936>). The data sets contain the results after DE testing with DESeq2 and count data, respectively. DE was determined by selecting genes with an adjusted  $P$ -value of <0.001. Recalibration was performed using the mean  $V^G_H$  values.

The Findley et al. (2021) data sets of different perturbation agents were downloaded from the supplemental data of the publication. The experiments mentioned in the analysis are "15C1" (copper perturbation) and "20C1" (zinc perturbation) of iPSCs from the shallow sequencing data set. The downloaded data sets contain the results of DE testing with DESeq2. DE was determined by selecting genes with a false discovery rate of <0.05. Recalibration was performed using the mean  $V^G_H$  values.

The psychENCODE data sets from Gandal et al. (2018) for ASD, SCZ, and BD were downloaded from the supplemental data. The data sets contain the results after DE testing with DESeq2. DE was determined using a false discovery rate cutoff of 0.05. Recalibration was performed using the inference-assisted  $V^G_H$  (i.e.,  $V^G_{HI}$ ) estimates for the tissue BRNCTXB (Brain Frontal Cortex BA9).

### Correlation between nominal and recalibrated fold changes

In order to avoid inflation caused by directionality (which does not change), we calculated the correlation between nominal and recalibrated

fold changes using the Pearson's method of the absolute values.

### GO term enrichment

Gene Ontology (GO) term enrichment analysis was conducted using the R package `g:Profiler2` (Raudvere et al. 2019; Kolberg et al. 2023), using GO terms and gene associations from the Biological Process (BP), Cellular Component (CC), and Molecular Function (MF) GO resources based on Ensembl release 110. Enrichments were calculated based on Ensembl gene identifiers against all genes with determined expression and  $V^G$  values in the data set. Genes were selected by absolute nominal fold change  $|FC|$  and absolute recalibrated fold change  $\overline{FC}$ , respectively. For all enrichment comparisons except the testing of other gene numbers, the top ~50% (rounded to the nearest 100) of all significant genes were selected. Only genes from the experiment for which  $V^G$  has been calculated were included as the background set. Enrichment was considered up to an adjusted  $P$ -value of 0.05. In the comparison of nominal and recalibrated fold changes, nonenriched GO terms were imputed with a  $P$ -value of 1.

GO term clustering to identify groups of GO terms with a high degree of shared genes was performed with the help of the R package `simplifyEnrichment` (Gu and Hübschmann 2023) using binary-cut clustering. Clustering had to be performed separately for terms of each GO resource (BP, CC, and MF).

### Gene set enrichment

Gene Set Enrichment Analysis (GSEA) was performed using the R package `clusterProfiler` (Wu et al. 2021) using `FGSEA` (Korotkevich et al. 2021) as backend. In contrast to GO term enrichment based on groups, only GO terms and gene association from the GO Biological Process resource were used, and the analysis was performed on all genes (irrespective of statistical significance) for which gene expression and  $V^G$  are defined. Enrichment was considered up to an adjusted  $P$ -value of 0.05. Corresponding nonenriched GO terms were imputed with a  $P$ -value of 1.

### Inferring $V^G$ estimates for additional tissues

The implemented inference method for gene  $V^G$  estimates per tissue comprises a series of subsequent steps, illustrated in Supplemental Figure S9 and added as Supplemental Code S2. The inference is based on the two  $26,760 \times 50$  matrices:  $E$  for gene expression (measured in transcripts per million) and  $V$  for  $V^G_H$ , with 26,760 being  $|G|$ , the number of genes  $G$  and 50 being  $|T|$ , the number of tissues  $T$ . Note that not all values in  $V$  are defined, as those are the estimates that are to be inferred. To circumvent issues with null values for very low-expressed genes,  $\bar{E}$  is defined by adding a pseudocount equal to the smallest expression value larger than 0 that is observed over all genes  $G$ :

$$\bar{E} = E + \min_{g \in G, t \in T, E_{g,t} > 0} E_{g,t}. \quad (2)$$

From these data, a model is fitted for each tissue  $t \in T$  over all genes  $g \in G$ . Accordingly, a linear least-square fit linking  $\log(\bar{E})$  and  $\log(V)$  is calculated for each tissue  $t$ , using the function "lm" in R:

$$\log(\bar{E}_t) = \beta_{0,t} + \beta_{1,t} * \log(V_t) + \epsilon. \quad (3)$$

Tissue  $V^G_H$  estimates are then adjusted using the fit to  $\hat{V}$ , the expression variance that would be expected if the tissue expression was equal to the mean expression over all tissues for

that gene:

$$\hat{V}_{g,t} = V_{g,t} * \left( \frac{1}{|T|} \sum_{r \in T} \bar{E}_{g,r} \right) / \bar{E}_{g,t} \Big)^{\beta_{1,t}} \quad (4)$$

The degree of similarity between the  $V_{g,t}^G$  estimates of different tissues  $t_1$  and  $t_2$  is determined using the Spearman's correlation coefficient  $\rho$  as

$$\rho(t_1, t_2) = \text{SpearmanCorrelation}(V_{t_1}, V_{t_2}). \quad (5)$$

For each gene  $g$  in each tissue  $t$ , we then define a set  $S_{g,t} \subset T$  consisting of the five tissues  $s \in S_{g,t}$  other than  $t$  that have the highest degree of similarity  $\rho(t, s)$  and for which  $V_{g,s}$  is defined. If  $V_g$  is defined for less than five tissues, then only those tissues are included. Unadjusted inferred  $V^G$  estimates  $\hat{V}$  are then calculated as a weighted mean of the estimates from the similar tissues, using  $\rho$  as weight:

$$\hat{V}_{g,t} = \frac{\sum_{s \in S_{g,t}} \rho(t, s) \cdot \hat{V}_{g,s}}{\sum_{s \in S_{g,t}} \rho(t, s)}. \quad (6)$$

As the final step,  $\hat{V}_{g,t}$  is adjusted to the expression value  $\bar{E}_{g,t}$  using the fit from Equation 3:

$$\tilde{V}_{g,t} = \left( \bar{E}_{g,t} / \left( \frac{1}{|T|} \sum_{r \in T} \bar{E}_{g,r} \right) \right)^{\beta_{1,t}} * \hat{V}_{g,t}. \quad (7)$$

We refer to  $\tilde{V}$  as inferred  $V^G$  estimates or  $V_{HI}^G$ . To increase the number of predicted genes, existing tissue estimates of  $V_{HI}^G$  were combined with  $V_{I}^G$ , with existing scores taking precedence if both existed for a particular gene, in what is called  $V_{HI}^G$ . All  $V_{HI}^G$  estimates are available in Supplemental Table S4.

### ASD genes

Previously ASD-associated genes were downloaded from An et al. (2018) using the most stringent list with  $FDR < 0.1$ . The tissue with the lowest  $V^G$  per gene was calculated from  $V_{HI}^G$ . GTEx brain tissues are all 13 GTEx tissues starting with "BRN" plus pituitary. Enrichment was determined using a one-sided binomial test.

### Data analysis

All data analysis was done in the R programming language (R Core Team 2021).

### Data access

All  $V_{HI}^G$  estimates are provided in Supplemental Table S3. All  $V_{HI}^G$  estimates are provided in Supplemental Table S4. An R package for recalibrating experimental DE data is available at GitHub (<https://github.com/LappalainenLab/recalibrate>). Version 1.0.0 of the recalibrate package is also provided as Supplemental Code.

### Competing interest statement

T.L. is an advisor to and owns equity in Variant Bio and has received speaker honoraria from AbbVie and Merck. The remaining authors declare no competing interests.

### Acknowledgments

We thank Sarah Kim-Hellmuth, Xiaoting Li, Ryan Collins, Sanna Gudmundsson, Paul Hoffman, Mariia Minaeva, and the current

and former members of the Lappalainen lab for helpful discussions and code sharing. Part of the computations were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at UPPMAX partially funded by the Swedish Research Council through grant agreement no. 2018-05973. This work was supported by funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 101043238) and the National Human Genome Research Institute of the National Institutes of Health (grant no. R01GM140287).

**Author contributions:** P.R., P.M., and T.L. conceived and designed the study. P.R., with assistance from A.K. and T.L., performed the data analysis. K.R.G. and P.M. contributed expertise regarding the  $V^G$  metric and generated the  $V_{HI}^G$  estimates. P.R. wrote the R package recalibrate. P.R. and T.L. wrote the manuscript. P.R., A.K., K.R.G., P.M., and T.L. revised and edited the manuscript.

### References

- Alasoo K, Rodrigues J, Mukhopadhyay S, Knights AJ, Mann AL, Kundu K, Hale C, Dougan G, Gaffney DJ. 2018. Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. *Nat Genet* **50**: 424–431. doi:10.1038/s41588-018-0046-7
- Albert FW, Kruglyak L. 2015. The role of regulatory variation in complex traits and disease. *Nat Rev Genet* **16**: 197–212. doi:10.1038/nrg3891
- An J-Y, Lin K, Zhu L, Werling DM, Dong S, Brand H, Wang HZ, Zhao X, Schwartz GB, Collins RL, et al. 2018. Genome-wide de novo risk score implicates promoter variation in autism spectrum disorder. *Science* **362**: eaat6576. doi:10.1126/science.aat6576
- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol* **57**: 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x
- Castel SE, Aguet F, Mohammadi P, Aguet F, Anand S, Ardlie KG, Gabriel S, Getz GA, Graubert A, Hadley K, et al. 2020. A vast resource of allelic expression data spanning human tissues. *Genome Biol* **21**: 234. doi:10.1186/s13059-020-02122-z
- Collins RL, Glessner JT, Porcu E, Lepamets M, Brandon R, Lauricella C, Han L, Morley T, Niestroj L-M, Ulirsch J, et al. 2022. A cross-disorder dosage sensitivity map of the human genome. *Cell* **185**: 3041–3055.e25. doi:10.1016/j.cell.2022.06.036
- Cui X, Churchill GA. 2003. Statistical tests for differential expression in cDNA microarray experiments. *Genome Biol* **4**: 210. doi:10.1186/gb-2003-4-4-210
- Domingo J, Minaeva M, Morris JA, Ghatan S, Ziosi M, Sanjana NE, Lappalainen T. 2024. Non-linear transcriptional responses to gradual modulation of transcription factor dosage. *eLife* **13**: RP100555. doi:10.7554/eLife.100555.1
- Dong D, Shen H, Wang Z, Liu J, Li Z, Li X. 2023. An RNA-informed dosage sensitivity map reflects the intrinsic functional nature of genes. *Am J Hum Genet* **110**: 1509–1521. doi:10.1016/j.ajhg.2023.08.002
- Findley AS, Monziani A, Richards AL, Rhodes K, Ward MC, Kalita CA, Alazizi A, Pazokitoroudi A, Sankararaman S, Wen X, et al. 2021. Functional dynamic genetic effects on gene regulation are specific to particular cell types and environmental conditions. *eLife* **10**: e67077. doi:10.7554/eLife.67077
- Fishilevich S, Nudel R, Rappaport N, Hadar R, Plaschkes I, Iny Stein T, Rosen N, Kohn A, Twik M, Safran M, et al. 2017. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database* **2017**: bax028. doi:10.1093/database/bax028
- Gandal MJ, Zhang P, Hadjimichael E, Walker RL, Chen C, Liu S, Won H, van Bakel H, Varghese M, Wang Y, et al. 2018. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**: eaat8127. doi:10.1126/science.aat8127
- Ge Y, Sealfon SC, Speed TP. 2009. Multiple testing and its applications to microarrays. *Stat Methods Med Res* **18**: 543–563. doi:10.1177/0962280209351899
- Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, Altman DG. 2016. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol* **31**: 337–350. doi:10.1007/s10654-016-0149-3
- The GTEx Consortium. 2020. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**: 1318–1330. doi:10.1126/science.aaz1776

- Gu Z, Hübschmann D. 2023. *SimplifyEnrichment*: a Bioconductor package for clustering and visualizing functional enrichment results. *Genomics Proteomics Bioinformatics* **21**: 190–202. doi:10.1016/j.gpb.2022.04.008
- Han X, Chen S, Flynn E, Wu S, Wintner D, Shen Y. 2018. Distinct epigenomic patterns are associated with haploinsufficiency and predict risk genes of developmental disorders. *Nat Commun* **9**: 2138. doi:10.1038/s41467-018-04552-7
- Harrison PF, Pattison AD, Powell DR, Beilharz TH. 2019. Topconfects: a package for confident effect sizes in differential expression analysis provides a more biologically useful ranked gene list. *Genome Biol* **20**: 67. doi:10.1186/s13059-019-1674-7
- Ignatiadis N, Klaus B, Zaugg JB, Huber W. 2016. Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. *Nat Methods* **13**: 577–580. doi:10.1038/nmeth.3885
- Jiang C-C, Lin L-S, Long S, Ke X-Y, Fukunaga K, Lu Y-M, Han F. 2022. Signalling pathways in autism spectrum disorder: mechanisms and therapeutic implications. *Signal Transduct Target Ther* **7**: 229. doi:10.1038/s41392-022-01081-0
- Jung K, Friede T, Beißbarth T. 2011. Reporting FDR analogous confidence intervals for the log fold change of differentially expressed genes. *BMC Bioinformatics* **12**: 288. doi:10.1186/1471-2105-12-288
- Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, et al. 2020. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**: 434–443. doi:10.1038/s41586-020-2308-7
- Kelly SA, Panhuis TM, Stoehr AM. 2012. Phenotypic plasticity: molecular mechanisms and adaptive significance. *Compr Physiol* **2**: 1417–1439. doi:10.1002/j.2040-4603.2012.tb00423.x
- Kim-Hellmuth S, Bechheim M, Pütz B, Mohammadi P, Nédélec Y, Giangreco N, Becker J, Kaiser V, Fricker N, Beier E, et al. 2017. Genetic regulatory effects modified by immune activation contribute to autoimmune disease associations. *Nat Commun* **8**: 266. doi:10.1038/s41467-017-00366-1
- Kolberg L, Raudvere U, Kuzmin I, Adler P, Vilo J, Peterson H. 2023. g:Profiler—interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Res* **51**: W207–W212. doi:10.1093/nar/gkad347
- Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, Sergushichev A. 2021. Fast gene set enrichment analysis. bioRxiv doi:10.1101/060012
- Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, et al. 2016. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**: 285–291. doi:10.1038/nature19057
- Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550. doi:10.1186/s13059-014-0550-8
- Manuck SB, McCaffery JM. 2014. Gene-environment interaction. *Annu Rev Psychol* **65**: 41–70. doi:10.1146/annurev-psych-010213-115100
- McCarthy DJ, Smyth GK. 2009. Testing significance relative to a fold-change threshold is a TREAT. *Bioinformatics* **25**: 765–771. doi:10.1093/bioinformatics/btp053
- Mohammadi P, Castel SE, Brown AA, Lappalainen T. 2017. Quantifying the regulatory effect size of *cis*-acting genetic variation using allelic fold change. *Genome Res* **27**: 1872–1884. doi:10.1101/gr.216747.116
- Mohammadi P, Castel SE, Cummings BB, Einson J, Sousa C, Hoffman P, Donkervoort S, Jiang Z, Mohassel P, Foley AR, et al. 2019. Genetic regulatory variation in populations informs transcriptome analysis in rare disease. *Science* **366**: 351–356. doi:10.1126/science.aay0256
- Mostafavi H, Spence JP, Naqvi S, Pritchard JK. 2023. Systematic differences in discovery of genetic effects on gene expression and complex traits. *Nat Genet* **55**: 1866–1875. doi:10.1038/s41588-023-01529-1
- Naqvi S, Kim S, Hoskens H, Matthews HS, Spritz RA, Klein OD, Hallgrímsson B, Swigut T, Claes P, Pritchard JK, et al. 2023. Precise modulation of transcription factor levels identifies features underlying dosage sensitivity. *Nat Genet* **55**: 841–851. doi:10.1038/s41588-023-01366-2
- Palmer D, Fabris F, Doherty A, Freitas AA, de Magalhães JP. 2021. Ageing transcriptome meta-analysis reveals similarities and differences between key mammalian tissues. *Aging* **13**: 3313–3341. doi:10.18632/aging.202648
- Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB. 2013. Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet* **9**: e1003709. doi:10.1371/journal.pgen.1003709
- Petrovski S, Gussow AB, Wang Q, Halvorsen M, Han Y, Weir WH, Allen AS, Goldstein DB. 2015. The intolerance of regulatory sequence to genetic variation predicts gene dosage sensitivity. *PLoS Genet* **11**: e1005492. doi:10.1371/journal.pgen.1005492
- Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, Vilo J. 2019. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res* **47**: W191–W198. doi:10.1093/nar/gkz369
- R Core Team. 2021. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. <https://www.R-project.org/>.
- Rice AM, McLysaght A. 2017. Dosage-sensitive genes in evolution and disease. *BMC Biol* **15**: 78. doi:10.1186/s12915-017-0418-y
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. *limma* powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res* **43**: e47. doi:10.1093/nar/gkv007
- Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140. doi:10.1093/bioinformatics/btp616
- Schena M, Shalon D, Davis RW, Brown PO. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**: 467–470. doi:10.1126/science.270.5235.467
- Sekovanić A, Jurasović J, Piasek M. 2020. *Metallothionein 2A* gene polymorphisms in relation to diseases and trace element levels in humans. *Arch Ind Hyg Toxicol* **71**: 27–47. doi:10.2478/aiht-2020-71-3349
- Starr AL, Gokhman D, Fraser HB. 2023. Accounting for *cis*-regulatory constraint prioritizes genes likely to affect species-specific traits. *Genome Biol* **24**: 11. doi:10.1186/s13059-023-02846-8
- Wang X, Goldstein DB. 2020. Enhancer domains predict gene pathogenicity and inform gene discovery in complex disease. *Am J Hum Genet* **106**: 215–233. doi:10.1016/j.ajhg.2020.01.012
- Wang H, Zhang F, Zeng J, Wu Y, Kemper KE, Xue A, Zhang M, Powell JE, Goddard ME, Wray NR, et al. 2019. Genotype-by-environment interactions inferred from genetic effects on phenotypic variability in the UK Biobank. *Sci Adv* **5**: eaaw3538. doi:10.1126/sciadv.aaw3538
- Wasserstein RL, Lazar NA. 2016. The ASA statement on *p*-values: context, process, and purpose. *Am Stat* **70**: 129–133. doi:10.1080/00031305.2016.1154108
- Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, et al. 2021. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation* **2**: 100141. doi:10.1016/j.xinn.2021.100141
- Zarrei M, MacDonald JR, Merico D, Scherer SW. 2015. A copy number variation map of the human genome. *Nat Rev Genet* **16**: 172–183. doi:10.1038/nrg3871
- Zhang S, Cao J. 2009. A close examination of double filtering with fold change and t test in microarray analysis. *BMC Bioinformatics* **10**: 402. doi:10.1186/1471-2105-10-402

Received January 9, 2025; accepted in revised form July 8, 2025.