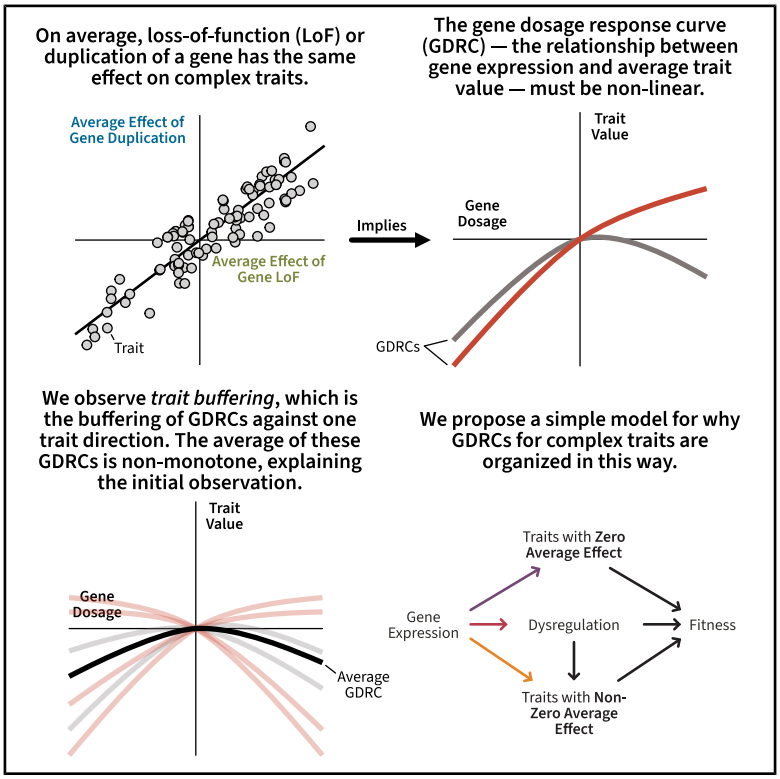
Cell Genomics

Article

Buffering of gene dosage response curves for human complex traits

Graphical abstract



Highlights

- Characterized gene dosage response curves (GDRCs) for 94 complex traits
- Analysis of dosage-altering variants shows that GDRCs are non-linear
- Across genes, expression changes tend to have weaker effects on one trait direction

Authors

Nikhil Milind, Courtney J. Smith, Huisheng Zhu, Tamara Gjorgjieva, Jeffrey P. Spence, Jonathan K. Pritchard

Correspondence

nmilind@stanford.edu (N.M.), jeff.spence@ucsf.edu (J.P.S.), pritch@stanford.edu (J.K.P.)

In brief

Milind et al. explore why loss-of-function variants and duplications tend to have average effects in the same direction on 94 complex traits. Using gene dosage response curves (GDRCs), they gather evidence for two explanatory models and interpret what these observations imply about the genetic architecture of complex traits.

Milind et al., 2026, Cell Genomics 6, 101221
July 8, 2026
https://doi.org/10.1016/j.xgen.2026.101221

CellPress

Cell Genomics

CellPress
OPEN ACCESS

Article

Buffering of gene dosage response curves for human complex traits

Nikhil Milind,[1,*] Courtney J. Smith,[1] Huisheng Zhu,[2] Tamara Gjorgjieva,[1] Jeffrey P. Spence,[1,3,4,*] and Jonathan K. Pritchard[1,2,5,*]

[1]Department of Genetics, Stanford University, Stanford, CA 94305, USA
[2]Department of Biology, Stanford University, Stanford, CA 94305, USA
[3]Institute for Human Genetics, University of California, San Francisco, CA 94143, USA
[4]Department of Epidemiology & Biostatistics, University of California, San Francisco, San Francisco, CA 94158, USA
[5]Lead contact
*Correspondence: nmilind@stanford.edu (N.M.), jeff.spence@ucsf.edu (J.P.S.), pritch@stanford.edu (J.K.P.)
https://doi.org/10.1016/j.xgen.2026.101221

SUMMARY

The genome-wide burdens of deletions, loss-of-function mutations, and duplications correlate with many traits. Curiously, for most of these traits, variants that decrease expression have the same genome-wide average direction of effect as variants that increase expression. This seemingly contradicts the intuition that for individual genes, reducing expression should have the opposite effect on a phenotype as increasing expression. To understand this paradox, we use the gene dosage response curve (GDRC), which relates changes in gene expression to expected changes in phenotype. We show that, for many traits, GDRCs are systematically biased in one trait direction relative to the other, a phenomenon we call trait buffering. Because of trait buffering, traits are more easily modified in one direction than the other by genetic variation. We develop a simple theoretical model that explains this bias in trait direction. Our results have broad implications for complex traits, drug discovery, and statistical genetics.

INTRODUCTION

Genome-wide association studies (GWASs) have identified thousands of variants associated with complex traits,[1,2] most of which are in non-coding regions of the genome.[3] This suggests that a substantial proportion of phenotypic variance is likely explained by variation in gene expression.[3–8]

One source of variation in gene expression is the gain or loss of functional copies of a gene.[9] Such gene dosage changes can have phenotypic consequences, with aneuploidies being an extreme example.[10–13] Copy-number variants (CNVs), which typically perturb the dosage of several genes, are another source of expression variation.[14] CNVs segregate in humans[15–18] and have measurable effects on various traits and disorders,[19–27] presumably by varying the expression of one or more of the genes with atypical copy numbers.[9]

There has been substantial work associating the total genome-wide CNV burden with traits.[19–22,24,25,27–30] An individual's genome-wide burden is estimated by counting the total number of CNVs they carry, regardless of which genes are affected. Associating this burden with a trait can be thought of as roughly estimating how much deleting or duplicating a random gene affects the trait. The large number of traits significantly associated with genome-wide deletion burden suggests that the effect of reducing the expression of a randomly chosen gene is often biased in a particular trait direction. Interestingly, genome-wide duplication burden and genome-wide deletion burden often have the same direction of effect (Methods S1, appendix A.2).[20–22] That is, increasing expression often has the same effect as decreasing expression on average. This seems to contradict the intuition that, for a given gene, increasing expression should have the opposite effect to decreasing expression.

To make sense of these genome-wide CNV burden test results from the perspective of what is happening at individual genes, we estimated the gene-level effects of loss-of-function (LoF) variants, deletions, and duplications using whole-exome sequencing (WES) and whole-genome sequencing (WGS) data from the UK Biobank (UKB) for 94 continuous traits.[31,32]

Our analysis revealed that across traits, the average effects of gene deletions and duplications are often non-zero and usually affect the trait in the same direction. The data also confirm the intuition that, for genes with large effects on traits, deletions generally have effects opposite to those of duplications. To explain these counterintuitive observations, we use the concept of gene dosage response curves (GDRCs), study their properties, and develop a simple model of their evolution.

RESULTS

Gene-level burden tests for LoF variants and duplications

To better understand how variants with different effects on dosage impact traits, we performed burden tests using LoF

variants, deletions, and duplications in the UKB. Our primary analyses focus on LoF variant burden tests because they are better powered than tests based on deletions, but deletion-based tests showed broadly similar patterns of effect (Methods S1, appendix I.5). We took care to keep the analysis of the different variant types as similar as possible so that the burden effect sizes could be interpreted jointly. We used the gene burden test implemented in REGENIE[33] and included covariates to correct for sex, age, batch effects, and other sources of confounding (STAR Methods). We also performed numerous quality control checks (Methods S1, appendices H.1 and H.4). We selected 410 continuous traits from various blood biochemistry, blood count, blood metabolite, and anthropometric measurements available in the UKB (data and code availability). We filtered these to a subset of 94 continuous traits with sufficient data and burden signal and used these for most of our analyses (STAR Methods). The summary statistics from these burden tests are provided as a resource to the community (data and code availability).

Connecting gene-level curves with genome-wide burden effects

Estimating the effect of genes on complex traits is central to understanding the mechanisms underlying trait biology. Changes in gene expression, which we refer to as "gene dosage," are often implicitly assumed to have a linear effect on a trait (Figure 1A). For example, transcriptome-wide association studies (TWASs) use a linear model for the relationship between predicted expression and trait value.[5,6] Under such a linear model, LoF and duplication burden estimates are expected to have opposite effects for any given gene (Figure 1A) because LoF variants and duplications likely affect traits by modulating gene dosage in *cis* (Methods S1, appendix I.5), which we observe empirically with protein levels (Methods S1, appendix A.5). The genome-wide burden effects for LoF variants and duplications, which can be thought of as averaging across these linear models for all genes, would be expected to have opposite signs. Thus, a linear model is incompatible with the genome-wide burden effects observed for various complex traits (Methods S1, appendix A.2) and suggests that these relationships must be, at a minimum, non-linear.

However, the interpretation of each genome-wide burden effect estimator at the gene level is complicated by the strategy used to define an individual's burden value. Previous studies have calculated various measures of genome-wide burden—counting the number of variants,[19–22,28] the total length of CNVs,[19,21,22,24,25,27] and the number of genes affected by CNVs.[22,24,27] We can approximately interpret these as the effect of perturbing a randomly chosen gene, assuming that the effect of CNVs occurs via perturbing individual genes, but this depends on other factors such as the local mutation rate of CNVs and the additivity of dosage effects across genes.

To estimate the actual effect of perturbing a randomly chosen gene, we estimated the effect of each individual gene on a trait by using burden tests and then averaged those estimates across all genes, which we call the average burden effect. This is in contrast to estimates of the genome-wide burden effect, which aggregate all CNV perturbations in the genome of an individual before testing their effect on a trait. We found that the estimated average burden effect for LoF variants was almost always in the same direction as that of duplications (Figure 1B), consistent with the genome-wide burden effects reported previously. Curiously, however, when we look at top hits from GWASs, we find no evidence for such an effect (Figure 1C; Methods S1, appendix I.1), consistent with previous analyses.[37]

Taken together, these observations provide contradictory information about how dosage is associated with traits. To explain these findings, we need to consider the gene-level heterogeneity present in dosage-response relationships. We envision the relationship between gene expression and trait as a GDRC: a continuous curve describing how changes in gene expression affect the average trait value. The core idea is that if a gene is involved in the biology of a trait, then different baseline expression values for the gene will, on average, result in different trait values. A given point on the GDRC is the hypothetical mean trait value of individuals whose gene expression is set to a particular level. In other words, the GDRC characterizes the relationship between expression and trait value for all possible dosage perturbations. The GDRC extends prior work exploring the quantitative relationship between gene dosage and trait value (e.g., allelic series[38] or experimental modulation[39–42]). Here, we propose that the GDRC acts as a conceptual framework to unify estimates of gene effects from LoF variants, duplications, and expression quantitative trait loci (eQTLs), allowing us to understand the source of non-linearity in the data.

Averaging over genetic backgrounds, a particular variant will have a specific effect on expression and thus correspond to a specific point on the GDRC. We define the origin to be the mean trait and expression value in the population (Methods S1, appendix A.4). A burden estimate of the LoF effect, $\widehat{\gamma}_{\mathrm{LoF}}$, is an estimate of a point on the left tail of the curve, while a burden estimate of the duplication effect, $\widehat{\gamma}_{\mathrm{Dup}}$, is an estimate of a point on the right tail of the curve (Figure 1A). The effect estimated by TWAS, $\widehat{\gamma}_{\mathrm{TWAS}}$, is the slope of a linear approximation to the GDRC in the region of expression levels spanned by common eQTLs. While we visualize LoF variants and duplications as having specific dosage effects on expression (50% and 150%, respectively), most of our analyses require only that they generally decrease or increase expression, respectively (Methods S1, appendix A.5).

Intuition suggests that GDRCs should be monotone, meaning that reducing and increasing expression should have opposite phenotypic effects (Figure 1D).[24,25,27,43,44] The average burden effects that we computed (Figure 1B) can be interpreted as points on the average GDRC (aGDRC) obtained by averaging the GDRC of each gene for a given trait. If average LoF and duplication effects are in the same direction, as we observe for many traits, it would imply a non-monotone aGDRC. It therefore seems surprising that aggregating presumably monotone GDRCs would result in a non-monotone aGDRC.

We suggest two models that could contribute to non-monotone aGDRCs. In the monotone model, monotone gene-level GDRCs that are systematically buffered against one trait direction could average to a non-monotone aGDRC (Figure 1E). In the non-monotone model, some GDRCs are non-monotone, and among the non-monotone GDRCs, it is more common to

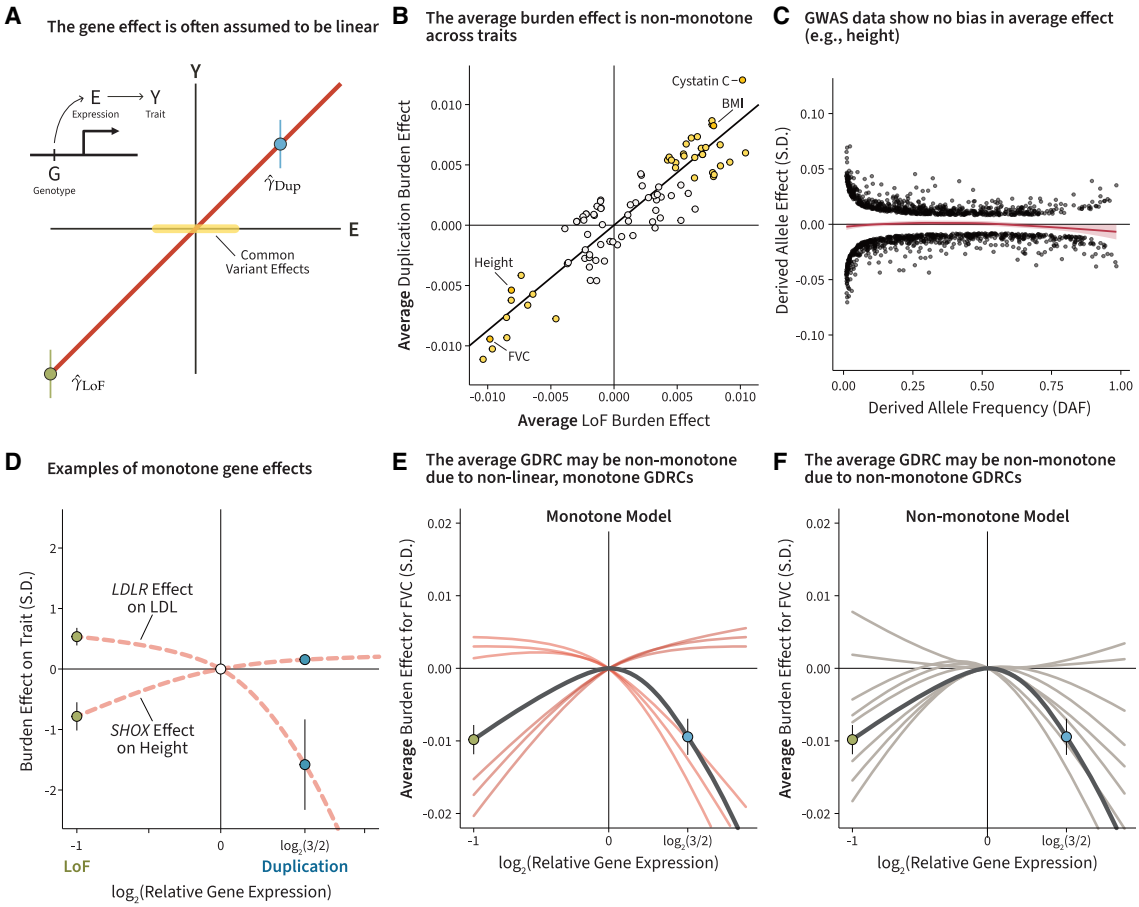


Figure 1. Explaining the paradoxical observations in burden data

(A) The hypothetical relationship between gene dosage (E) and expected trait value (Y) is often assumed to be linear. LoF variants and duplications estimate the gene effect at different dosage values ($\hat{\gamma}_{LoF}$ and $\hat{\gamma}_{Dup}$).

(B) The average burden effect from LoF variants and duplications across traits. Yellow points represent significant effects from both variant classes ($p < 0.05$), and the regression line was estimated using total least squares ($\beta = 0.87$, $z = 20.42$).

(C) Conditionally independent GWAS top hits for height. The locally estimated scatterplot smoothing (LOESS) curve shows no bias across the frequency spectrum. For variants with a derived allele frequency less than 0.5, the proportion of trait-increasing alleles is 51% (95% confidence interval [CI]: [0.48, 0.55]). For further discussion, refer to Methods S1 (appendix I.2).

(D) Burden estimates and 95% confidence intervals for two illustrative gene-trait pairs, with hypothetical GDRCs as dashed lines. *LDLR* encodes the low density lipoprotein (LDL) receptor, a known negative regulator of LDL levels in blood, and acts by sequestering LDL out of the bloodstream.[34,35] *SHOX* haploinsufficiency is associated with short stature, while increased dosage is associated with tall stature.[36]

(E) The first model explains the non-monotone average GDRC of forced vital capacity (FVC) (with 95% CIs) using monotone gene curves.

(F) The second model explains the non-monotone average GDRC of FVC (with 95% CIs) using non-monotone gene curves.

BMI, body mass index.

be non-monotone in a particular direction, driving the aGDRC to be non-monotone (Figure 1F). These models are not mutually exclusive and may both contribute to the non-monotone aGDRC observed for a trait.

Top hits are consistent with monotone GDRCs

To understand the relative contribution of these models to aGDRCs, we started by looking at the results from the burden tests. Using a Bayesian analysis of the p value distribution,[45] we conservatively estimate that 5% of the gene-trait pairs have a true association (Methods S1, appendix A.6). This suggests that even at current sample sizes, burden tests of LoF variants and duplications are relatively underpowered and that it will be challenging to explain the non-monotone aGDRCs using only ascertained GDRCs.

We confirmed this intuition by performing an exome-wide analysis. We performed a stringent multiple testing correction on our burden tests, accounting for the number of genes, number of traits, and both types of tests conducted (Bonferroni p value threshold of $p < 2.87 \times 10^{-8}$, equivalent to $|z| > 5.55$). At this threshold, we detected 443 LoF hits and 677 duplication hits. However, these hits were mostly non-overlapping. Only two hits overlapped, both of which indicated monotone GDRCs (Figure 2).

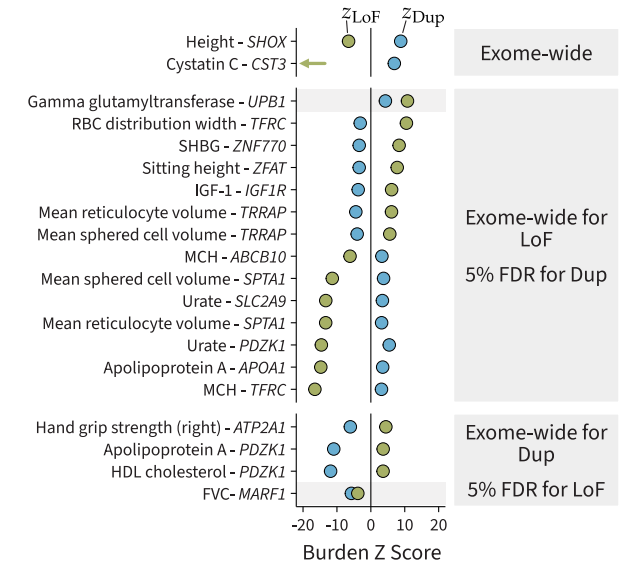


Figure 2. Top hits from burden tests

The Z scores from regression-based burden tests using LoF and duplication variants across various continuous traits. Only two genes had exome-wide significant LoF and duplication burden hits. To reduce the multiple-testing burden, we performed exome-wide ascertainment of LoF burden tests and ascertained the duplications for the subset at a 5% false discovery rate (FDR). We performed the reciprocal test using the duplications. The green arrow represents a Z score smaller than -20. Non-monotone relationships are highlighted in gray.

SHBG, sex hormone binding globulin; IGF-1, insulin-like growth factor 1; MCH, mean corpuscular hemoglobin; HDL, high density lipoprotein; FVC, forced vital capacity.

To reduce the multiple testing burden, we used the 443 LoF hits as a discovery set and ascertained genes with significant duplication burden tests at a 5% false discovery rate (FDR). We also performed the reciprocal analysis using the duplication burden tests as a discovery set. Consistent with intuition, most top genes have monotone effects on traits (Figure 2), although we do observe two non-monotone GDRCs. At a 5% FDR, we expect 0.5 of the 20 discoveries to be a non-monotone false discovery, providing some evidence for their existence. However, it is difficult to know their relative proportions, let alone how much they contribute to the non-monotone aGDRCs, so we next developed quantitative methods to study monotonicity.

Burden signal for traits is largely monotone

Due to the sparsity of the burden signal, we focused on quantitative measures that average signals across genes to explain the non-monotone aGDRCs (Figures 1E and 1F). First, we wanted to develop a notion of whether GDRCs generally tend to be monotone or not, beyond just considering the handful of significant hits.

We started with the observation that monotone GDRCs imply that LoF variants and duplications should have effects in opposite directions (Figure 3A), which we visualized by plotting the LoF burden effect against the duplication burden effect. If a duplication and an LoF variant have a large effect in the same direction, the GDRC must be non-monotone, and the product of their effects would be positive. If the GDRC is monotone, then LoF variants and duplications will have opposite effects, and their product will be negative. We define ϕ to be a normalized version of the negative product of the effect sizes (STAR Methods; Methods S1, appendix B.2). ϕ is positive if genes with large effects on a trait tend to have monotone effects on average. Note that ϕ does not directly measure the proportion of monotone GDRCs, and its interpretation requires some care (Methods S1, appendix F).

We used a maximum likelihood approach to estimate ϕ. We performed extensive simulations from the individual-level genotype data to validate that our estimator, $\widehat{\phi}$, is relatively unbiased and well calibrated (Methods S1, appendix G.1.5). Across traits, we found that the estimated values of ϕ were generally positive (Figure 3B), suggesting that the burden signal is explained predominantly by monotone GDRCs. The same trend toward monotone signal was observed using a method-of-moments estimator for ϕ that makes fewer distributional assumptions (Methods S1, appendix I.4). This agrees with our biological expectations of monotone gene-level effects and is potentially consistent with the monotone model.

A model for how non-monotone GDRCs arise

Our estimates of ϕ suggested that monotone GDRCs play a significant role in complex traits, but even among the small number of significant hits for individual gene-trait pairs, we detected two non-monotone associations (Figure 2). While averaging across all genes indicates that the predominant signal is generally monotone (Figure 3), this does not preclude the possibility that non-monotone GDRCs, in aggregate, contribute to the non-monotone aGDRCs. It is simultaneously possible to have positive values of ϕ and significant contributions from non-monotone GDRCs to the aGDRC (Methods S1, appendix F).

Indeed, prior evidence suggests that non-monotone GDRCs play some role in complex traits. One line of evidence comes from the analysis of genomic disorders caused by reciprocal CNVs, which are loci where both deletions and duplications are observed. Many reciprocal CNVs are known to induce the same phenotype,[46,47] suggesting that the effect of increasing or decreasing the dosage of the disorder-associated gene has the same direction of effect on the trait. Another line of evidence comes from careful analysis of CNVs in the UKB, where multiple loci show evidence of non-monotone associations with various complex traits.[27,48] This suggests that, although challenging to detect at an exome-wide level, non-monotone GDRCs may still contribute to the non-monotone aGDRCs.

Before investigating genome-wide signals for non-monotone GDRCs, we wanted to develop an intuition for how they might arise. Consider a gene that affects a trait via a single biological pathway (Figure 4A). Intuition about biological mechanisms suggests that each step along this pathway should be monotone but possibly non-linear. For instance, protein levels are often buffered against large changes in gene expression,[9] which results in a non-linear but monotone response of protein levels to gene expression. Similarly, perturbations of the

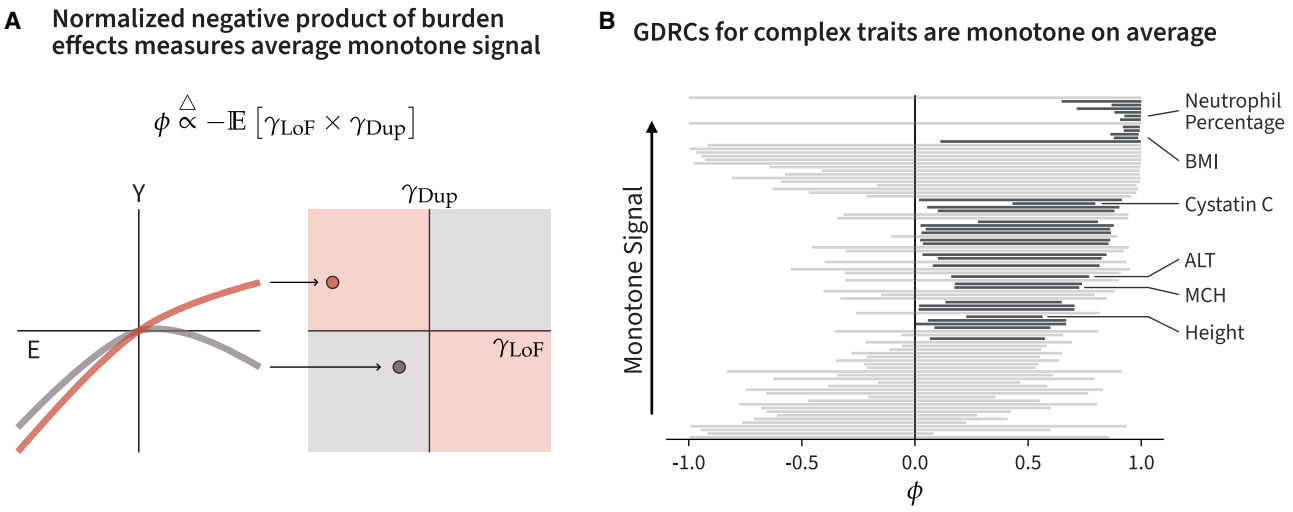


Figure 3. Monotonicity inference for traits

(A) Monotone curves (red) and non-monotone curves (gray) fall in different quadrants on a plot of the duplication burden effect versus the LoF burden effect. A natural estimate to compare points in these quadrants is the negative product of the effect sizes, which is positive when the curve is in the red quadrants and negative when it is in the gray quadrants. We define ϕ as the normalized version of this product, averaged across all curves.

(B) The bars represent 95% confidence intervals for ϕ for the traits in our analysis. The black bars are significantly non-zero ($p < 0.05$). The bars are ordered by $\widehat{\phi}$. BMI, body mass index; ALT, alanine transaminase; MCH, mean corpuscular hemoglobin.

expression levels of transcription factors can result in non-linear but mostly monotone changes in the expression of their targets.[40,41] As the dosage effect percolates through a pathway toward the focal trait, these curves compose with one another to form the relationship between gene dosage and trait. If each curve along the pathway is monotone, then the overall relationship between gene expression and trait is mathematically guaranteed to also be monotone (Figure 4A; Methods S1, appendix A.7).

One example of a monotone association in our data is that of *SLC2A9* expression and urate levels (Figure 2). *SLC2A9* encodes a transporter expressed in the kidneys that is responsible for reabsorbing urate into the blood by transporting urate from the kidney into the bloodstream.[49,50] A simple model of the observed monotone GDRC is that *SLC2A9* expression has a monotone effect on urate reabsorption, which in turn has a monotone effect on urate levels in blood (Figure 4A).[51] The composition of these two monotone steps can explain the monotone GDRC we observe in the data.

However, non-monotone GDRCs can arise if a gene's effects flow through two or more pathways (Figure 4B).[52] We propose a simple, hypothetical model for the non-monotone relationship between *UPB1* expression and gamma-glutamyl transferase (GGT) levels in blood (Figure 2). Both LoF variants and duplications in *UPB1* are associated with an increase in GGT levels (Figure 4B). GGT is an enzyme found in the membrane of epithelial cells in various tissues, particularly in the hepatobiliary system, and acts as a readout of oxidative stress.[53–55] *UPB1* encodes an enzyme involved in the breakdown of pyrimidines in the liver to downstream byproducts such as beta-alanine. Both excess pyrimidine metabolites and excess beta-alanine can result in toxicity and increased oxidative stress.[56–58] Thus, we propose that increased *UPB1* expression has two monotone effects: it decreases the concentration of pyrimidines and increases the concentration of beta-alanine, both of which can contribute to increased GGT levels by increasing oxidative stress. The total effect of these two pathways could explain the non-monotone GDRC observed between *UPB1* and GGT levels (Figure 4B). Generalizing these biological principles, non-monotone GDRCs can arise whenever the effect of a gene propagates through multiple pathways to affect the trait (Figure 4B).

Genes are buffered against one trait direction

One interpretation of the GDRCs in the monotone and non-monotone models (Figures 1E and 1F) is that traits are more likely to be perturbed in one direction than the other, regardless of the direction of expression perturbation. Therefore, this is a form of buffering that occurs at the trait level, which we refer to as trait buffering. We endeavored to quantify how much monotone and non-monotone GDRCs contributed to trait buffering and non-monotone aGDRCs.

When plotting duplication burden effects against LoF burden effects, genes experiencing trait buffering preferentially map to a region that is away from a diagonal line (Figure 5A). The diagonal line represents perfectly linear GDRCs (Methods S1, appendix D.1). Points above the diagonal line correspond to GDRCs that are buffered against negative trait values, and points below the diagonal line correspond to GDRCs buffered against positive trait values.

To develop some preliminary intuition about the trait buffering model, we estimated γ_{LoF} and γ_{Dup} for each gene. Since burden effect estimates are noisy for individual genes, we performed Bayesian inference using a flexible multivariate adaptive shrinkage (MASH) prior,[59] which models the joint distribution of γ_{LoF} and γ_{Dup} as a mixture of bivariate normal distributions that is learned from the data. We used posterior samples of γ_{LoF}

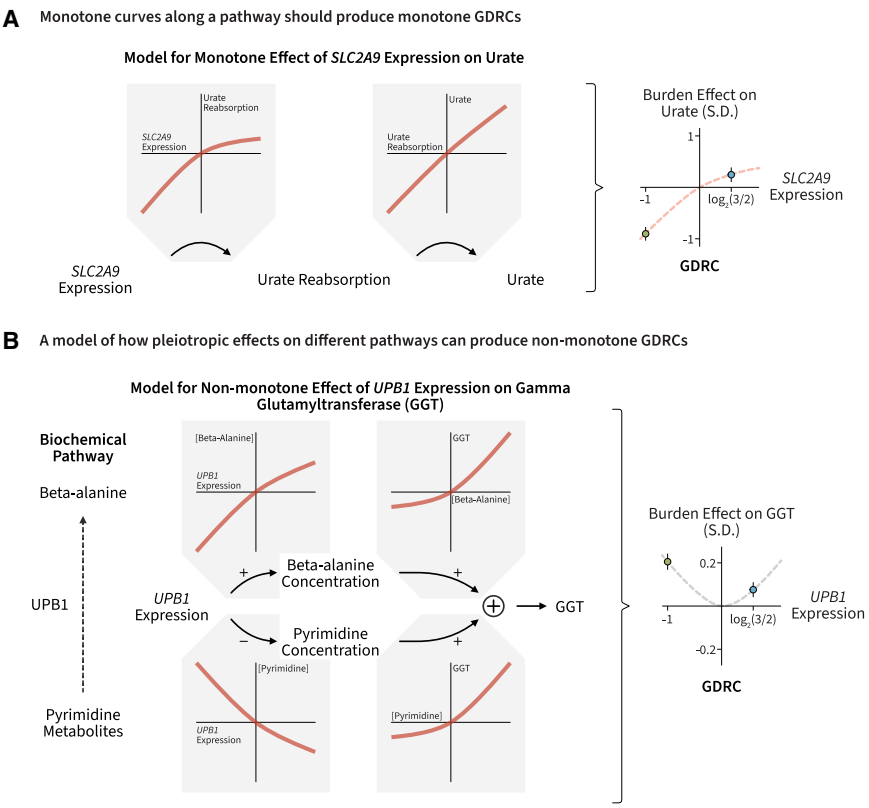


Figure 4. Model for non-monotone GDRCs

(A) A simple model on the left, composed of monotone curves at each step, can explain the monotone GDRC observed for *SLC2A9* and urate levels. *SLC2A9* encodes a kidney transporter responsible for reabsorbing urate back into the blood. Thus, increased *SLC2A9* expression may contribute to increased urate reabsorption, which in turn can increase the levels of urate in blood.

(B) In our model, the effect of *UPB1* propagates through two separate paths to affect gamma-glutamyl transferase (GGT) levels, which may explain the non-monotone GDRC. The first path, similar to our monotone example, consists of two increasing steps (denoted by + symbols above the arrows). The second path consists of one increasing and one decreasing step (denoted by the – symbol below the arrow). The outcomes of the paths are summed in our conceptual model, resulting in a non-monotonicity that might explain the observed data.

and γ_{Dup} for each gene to better understand how genes might contribute to trait buffering. As an example, the posterior means of γ_{LoF} and γ_{Dup} for height are displayed in Figure 5B, and further examples are provided in Methods S1 (appendix I.7). Both monotone and non-monotone GDRCs are predominantly below the diagonal, concordant with the direction of the aGDRC for height.

To test this more formally, we developed a statistic, ξ, to measure trait buffering. ξ aggregates the signed strength of deviation from the diagonal line across all GDRCs (Methods S1, appendix D). We calculate the statistic using posterior samples of γ_{LoF} and γ_{Dup}.

The sign of ξ indicates which trait direction is being buffered against. For example, based on their aGDRCs, we would predict that height should have a negative ξ value and cystatin C should have a positive ξ value. If GDRCs are not preferentially buffered against either direction of a trait, our statistic should be close to zero. Indeed, this measure of trait buffering showed strong concordance with the direction of the non-monotone aGDRCs (Methods S1, appendix I.8).

To understand the extent to which our proposed models (Figures 1E and 1F) contributed to trait buffering, we decomposed ξ into the contribution from non-monotone and monotone GDRCs (STAR Methods; Methods S1, appendix D) and estimated these components from genes for which we were relatively confident in the sign of effect using a local false sign rate filter. Note that this decomposition does not quantify the proportion of monotone and non-monotone GDRCs but rather the extent to which these curves contribute to the observed aGDRCs (Methods S1, appendix F). These measures revealed that both monotone and non-monotone curves contribute to trait buffering (Figure 5C). These patterns persist as we increase the confidence in the sign of the gene's effect (Figure 5C). We performed extensive simulations to demonstrate how these component estimates behave under various extreme scenarios (Methods S1, appendix G.2). Juxtaposed with these simulations, our results for certain traits (e.g., forced vital capacity [FVC], peak expiratory flow [PEF], forced expiratory volume in 1 second [FEV1], body mass index [BMI], and height) show a clear signature of non-monotone genes having a substantial contribution to trait buffering. Ultimately, genes with both monotone and non-monotone GDRCs (Figures 1E and 1F) likely contribute to observed non-monotone aGDRCs.

A dysregulation phenotype may be upstream of many complex traits

To recapitulate, we found evidence for two models contributing to the observed average burden effects. These models imply that GDRCs for complex traits are organized in a manner that results in trait buffering.

Two observations initially seem to imply that the complex traits with non-monotone aGDRCs are under directional selection. First, the GDRCs for these traits are buffered against a particular trait direction. Second, non-monotone aGDRCs suggest that any variant should, on average, have a non-zero effect in the same direction, regardless of whether the variant increases or decreases expression.

To see why these observations are signatures of directional selection, we consider how the average expression of a given gene will evolve over time in response to directional selection. Suppose a gene affects a trait where higher trait values are

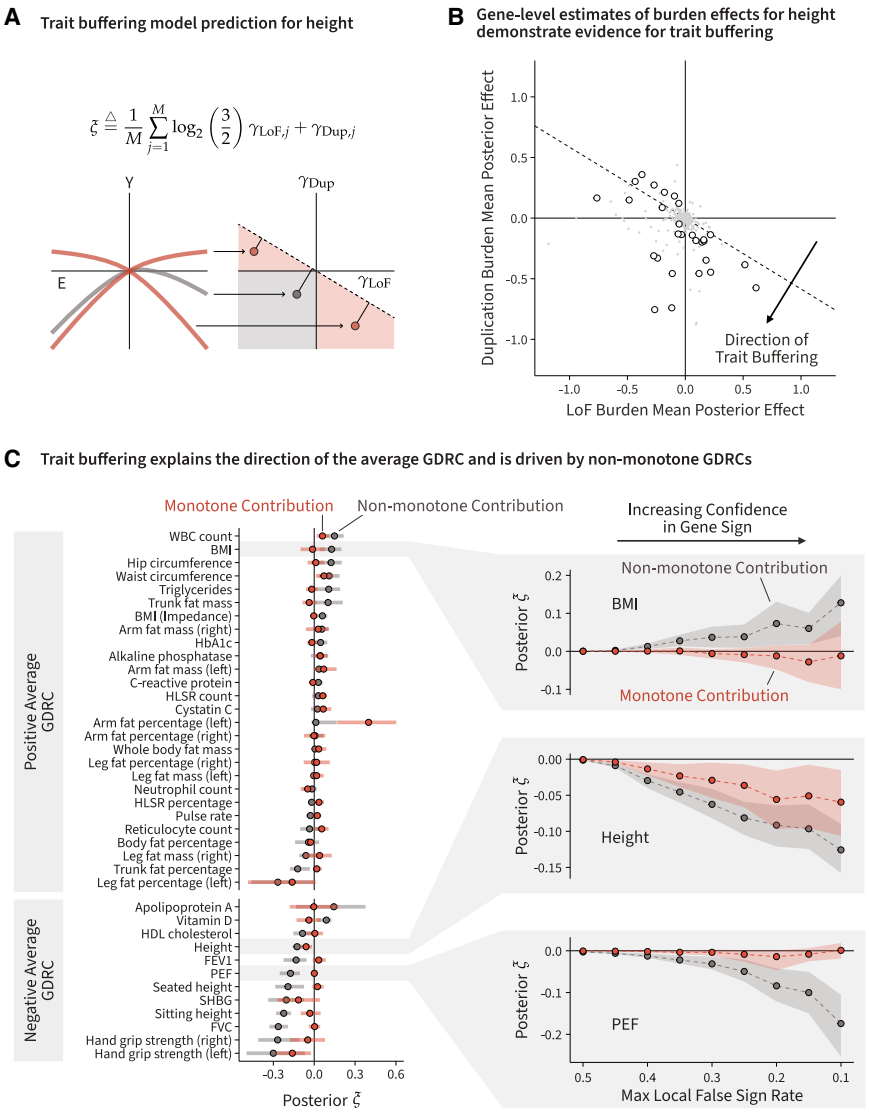


Figure 5. Trait buffering as a model for GDRC architecture

(A) After combining both models of GDRCs contributing to the non-monotone aGDRC, the GDRCs fall below a diagonal line for traits with negative average burden effects, such as height. The average LoF and duplication burden effects will be in the direction of trait buffering. ξ measures the average signed deviation from the diagonal line, represented by the lines orthogonal to the diagonal line.

(B) Estimates of the latent burden effects show, visually, that height is consistent with trait buffering. Each point represents a gene. The large circles represent genes for which the local false sign rate for both LoF and duplication burden effects is less than 10%.

(C) (Left) The posterior estimates of trait buffering, ξ, decomposed into contributions from non-monotone and monotone GDRCs. These estimates are derived from genes with a local false sign rate of less than 10% for LoF and duplication burden effects. (Right) The estimates of the components remain non-zero with increasing confidence in the signs of the gene-level effects.

WBC, white blood cell; BMI, body mass index; HLSR, high light scatter reticulocyte; HDL, high density lipoprotein; FEV1, forced expiratory volume in 1 second; PEF, peak expiratory flow; SHBG, sex hormone binding globulin; FVC, forced vital capacity.

selected against (Figure 6A). Then, selection will tune the expression of the gene until the trait is minimized, resulting in a non-monotone GDRC where both deletions and duplications increase the trait. Thus, the non-monotone aGDRC and trait buffering in the direction of the average burden effects suggest that the GDRCs of these complex traits may be shaped by directional selection.

However, prior analyses using GWAS data for many of these traits suggest that they are under stabilizing selection, not directional selection.[60–64] Additionally, we do not observe a non-zero average effect in GWAS data (Figure 1C; Methods S1, appendix I.2), inconsistent with directional selection.

Stabilizing selection affects GDRCs differently from directional selection. Suppose a gene affects a trait under stabilizing selection (Figure 6B). Stabilizing selection acts to maintain an optimum trait value and will tune the expression of the gene to reach the optimum trait value. Unlike in the case of directional selection, the local shape of the GDRC around the optimum trait value is not restricted to minima. In this case, we do not expect any particular GDRC shape to dominate, nor do we expect LoF variants or duplications to be biased in a particular direction.

To explain the incongruous non-zero average burden effects and zero average GWAS effect, we propose that traits with non-monotone aGDRCs are downstream of an unobserved trait under directional selection that we refer to as "dysregulation" (Figure 6C). If the focal trait is downstream of dysregulation, the GDRC of the gene for the focal trait may be shaped by both forms of selection, depending on whether the gene is a direct or indirect regulator of the focal trait (Figure 6D). For genes that act primarily on the trait via dysregulation, the GDRCs will be non-monotone and in the direction of increasing dysregulation. For genes that directly modulate the focal trait, we expect no particular preference for monotone or non-monotone GDRCs. Genes that act on the focal trait both directly and through dysregulation explain the monotone curves that experience trait buffering.

Overall, this model explains the non-monotone average effects that we observe in the data. Depending on the balance between direct and indirect regulation via dysregulation, we can still have predominantly monotone GDRCs but a consistent bias toward one direction of the focal trait.

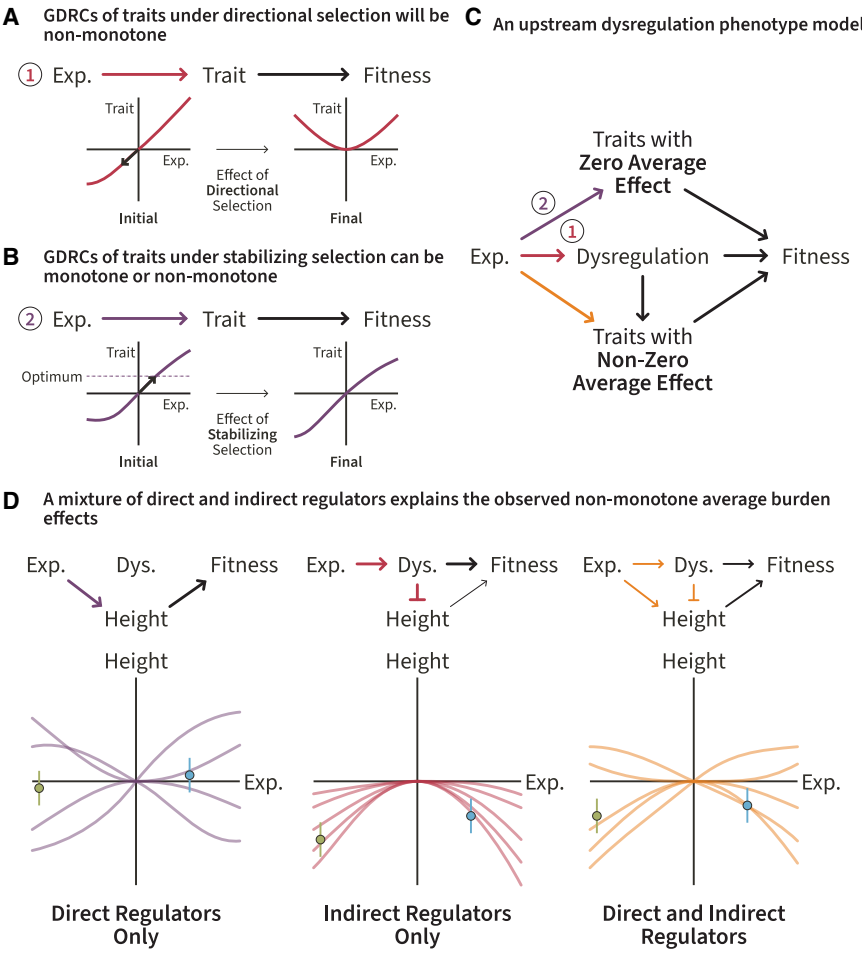


Figure 6. A theoretical model to explain trait buffering

(A) GDRCs of traits under directional selection will become non-monotone over evolutionary time.

(B) GDRCs of traits under stabilizing selection will have no preference for GDRC shape.

(C) In our model, traits with non-monotone aGDRCs have an upstream dysregulation phenotype, which we hypothesize is under directional selection.

(D) Example GDRCs of genes that follow specific paths to affect the focal trait (for example, height) under our theoretical model.

Exp., expression; Dys., dysregulation.

Our model also explains why the non-zero average effect is apparent in LoF variants and duplications across various traits but is not discernible in GWAS effect sizes. Consider two variants that have the exact same effect on the focal trait. One variant affects the trait directly, while the other affects the trait via dysregulation. Under our model, the second variant, which acts via dysregulation, can affect the focal trait in only one direction because it has a non-monotone GDRC, whereas the first variant has no such restriction. Both variants will experience the exact same fitness consequences from stabilizing selection on the trait. However, the variant acting via dysregulation will experience additional fitness consequences. Thus, holding all other factors equal, the variant acting via dysregulation will have a lower allele frequency and be more challenging to detect in a GWAS, thus making it less likely to contribute to a non-zero average effect.

We call this hypothetical latent phenotype "dysregulation" because the downstream traits with large non-monotone average burden effects are readouts of general dysfunction. The traits with significant non-monotone average effects (Methods S1, appendix I.1) include measures of organ dysfunction such as alkaline phosphatase, cystatin C, and C-reactive protein, which all have positive average LoF and duplication effects. Put another way, both deletion and duplication of a random gene tend to increase disease risk and measures of organ dysfunction. Similarly, measures of lung capacity and muscle strength have negative average LoF and duplication effects. Overall, the average effect is often in the direction of dysregulation across the continuous traits we studied.

Recent analyses of predicted biological age in the UKB also support the presence of a dysregulation phenotype. Age is associated with general dysregulation and disease.[65] Strong predictors of biological age in published predictive models include FEV1, cystatin C, HbA1c, alkaline phosphatase, and hand grip strength,[66,67] suggesting that these continuous traits with large non-monotone average effects may act as readouts of dysregulation.

DISCUSSION

In this study, we interpreted dosage-perturbing rare variants through the lens of the GDRC. Since we currently lack the power to analyze these data at the gene level, we designed quantitative measures to understand the average behavior of GDRCs across genes. Although we did not quantify the proportion of monotone or non-monotone GDRCs, our analysis suggests that both monotone and non-monotone GDRCs contribute to the concordant average burden effects of LoF and duplication variants. This results in GDRCs that are buffered against one trait direction. We hypothesized that this may be explained by the effect of genes on an upstream dysregulation phenotype, resulting in directional selection shaping GDRCs.

Limitations of the study

While we can estimate genome-wide patterns (i.e., average burden effect, monotonicity, and trait buffering) with high confidence, we are underpowered to estimate the properties of individual genes. In particular, while we can estimate that most of the

burden signal for traits is monotone, we are unable to get a precise estimate for the proportion of genes that are non-monotone. In addition, we can estimate only three points on GDRCs using the data generated in this paper, limiting the types of characteristics that we can infer about these curves.

Implications of the observed characteristics of GDRCs

In a ''linear world'' where dosage has a linear relationship with a trait, ascertaining any point on the GDRC along the dosage spectrum would carry all the information about the gene's effect on the trait. However, biology is inherently non-linear, and our analysis shows that these assumptions break down for variants with large effects. This has significant implications for drug design. If the GDRC of the target gene is non-monotone, perturbation in either direction may only increase disease risk. This occasional lack of concordance between LoF and duplication burden effects motivates the use of assays that test both ends of the dosage spectrum.[68] The shape of the GDRC implied by the burden estimates can also be a useful prior for fine-mapping causal variants[69] and linking variants to genes.[70–74]

The assumption of linearity in TWAS[5,6,75] and Mendelian randomization[76] approaches may cause estimated effect sizes to be attenuated toward zero for non-monotone genes. The ability of TWAS to recover trait-associated genes with non-linear GDRCs needs to be explored further. Yet TWAS-type methods would be invaluable for understanding the behavior of the GDRC in the physiological range of expression, thus motivating the continued development of such methods.

Toward the inference of gene-level curves across tissues and contexts

Moving from genetic associations toward mechanistic insights for complex traits remains challenging. The GDRC captures a large fraction of the biological insights that we want to extract from association studies. Although inference of the GDRC may be challenging, it is worth considering which aspects of the GDRC we can estimate using population-level and experimental approaches.

In principle, each tissue or cell type has its own specific GDRC for a gene-trait pair. In our approach, we focused on the global GDRC—which we could interpret as a weighted sum of tissue-specific GDRCs—by using variants that affect all tissues and contexts to understand the aggregate effect of dosage perturbation on traits. Modern approaches that estimate the effect of genes in specific tissues or cell types, such as TWASs, can be thought of as estimating or approximating aspects of tissue-specific GDRCs.

Even with methodological advances, it is unlikely that we will be able to infer all points of the GDRC of a gene with population-level data alone. The effects of selection make it challenging to detect variants in genes with large effects on traits.[77,78] This suggests that the ability to effectively infer the GDRC will vary across the genome and likely be most challenging for genes that are particularly important to the biology of traits.

These limitations of population-level data suggest that sequence-to-expression models[79–82] and experimental approaches[83–85] will play a critical role in fleshing out the GDRC for key genes. Missense variants, which in aggregate have a measurable effect on complex traits,[31] are challenging to integrate into the GDRC framework. Accurate computational effect prediction[86] or experimental approaches such as deep mutational scanning[87] are critical for translating the effect of missense variants into an ''effective dosage,'' which would allow them to be placed on the GDRC. Similarly, for common variants, we are restricted to using dosage effect estimates from eQTL studies in tissues that have been assayed, which are often post-mortem and from adult donors.[88] However, we cannot feasibly assay all tissues and cell types in all contexts. Thus, predicting common variant effects from sequence and multi-omic context or using experimental approaches such as massively parallel reporter assays[84] will be necessary to estimate full GDRCs. New techniques are also being developed to modulate dosage directly to measure molecular or cellular phenotypes,[41,68,89–92] which can provide *in vitro* measurements of GDRCs.

In summary, we have explored the nature of the relationship between gene dosage and trait for various complex traits. Our final model explains the initial observation of non-monotone average effects and provides intuition for how these relationships may be shaped by natural selection. The insights about GDRCs underlying complex traits will be informative for various downstream applications, including therapeutics and methods development.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Jonathan K. Pritchard (pritch@stanford.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- All genetic and health data were acquired from the UKB, a biomedical database containing information from half a million UK participants (https://www.ukbiobank.ac.uk/). Data for genetic association analysis of continuous traits were acquired under application 52374. Data for genetic association analysis of metabolite traits were acquired under application 30418. These data are available upon application to the UKB. All summary statistics generated from genetic association analyses and other processed data files are deposited in Zenodo: https://www.doi.org/10.5281/zenodo.19140003.
- Scripts used to process and analyze data are deposited in Zenodo: https://www.doi.org/10.5281/zenodo.19140111. Scripts were executed either on the UKB Research Analysis Platform (https://ukbiobank.dnanexus.com/) or on the Sherlock High-Performance Computing Cluster managed by the Stanford Research Computing Center (https://www.sherlock.stanford.edu/).

ACKNOWLEDGMENTS

We thank the members of the Pritchard lab for their feedback on this project and manuscript. In addition, we thank Yi Ding and Zeyun Lu (Gusev Lab) for reading over and providing feedback on the initial draft. We thank Luke O'Connor, Craig Smail, and an anonymous reviewer for their careful reading and feedback on our initial submission, which helped improve the paper during peer review. This research has been conducted using data from the UK Biobank (https://www.ukbiobank.ac.uk/), a major biomedical database. We thank all participants and researchers involved in the UK Biobank for their contribution and Nightingale Health PLC for early access to the NMR metabolomics data. Computing for this project was performed on the Sherlock

High-Performance Computing Cluster. We thank Stanford University and the Stanford Research Computing Center for providing computational resources and support that contributed to this research. N.M. was supported by a National Science Foundation Graduate Research Fellowship and Stanford's Knight-Hennessy Scholars Program. C.J.S. was supported by Stanford's Knight-Hennessy Scholars Program and the Stanford Center for Computational, Evolutionary and Human Genomics. H.Z. was supported by the Stanford Biology Department's graduate student assistantship. T.G. was supported by Stanford's Knight-Hennessy Scholars Program. This work was supported by the National Institutes of Health (R01HG011432, R01HG008140, and R01HG014005 to J.K.P.).

AUTHOR CONTRIBUTIONS

Methodology, N.M., C.J.S., H.Z., T.G., J.P.S., and J.K.P.; software, N.M., C.J.S., and T.G.; formal analysis, N.M.; data curation, N.M.; writing – original draft, N.M., C.J.S., H.Z., T.G., J.P.S., and J.K.P.; visualization, N.M., C.J.S., and H.Z.; conceptualization, J.P.S. and J.K.P.; supervision, J.P.S. and J.K.P.; project administration, J.K.P.; funding acquisition, J.K.P.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

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

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Inclusion and ethics
- METHOD DETAILS
  - Phenotypes
  - Genotypes
  - Burden tests
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Statistics and reproducibility
  - Monotonicity model
  - Trait buffering model
  - Data analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xgen.2026.101221.

Received: January 6, 2026
Revised: February 27, 2026
Accepted: April 1, 2026
Published: May 7, 2026

REFERENCES

1. Claussnitzer, M., Cho, J.H., Collins, R., Cox, N.J., Dermitzakis, E.T., Hurles, M.E., Kathiresan, S., Kenny, E.E., Lindgren, C.M., MacArthur, D.G., et al. (2020). A brief history of human disease genetics. Nature *577*, 179–189.
2. Sollis, E., Mosaku, A., Abid, A., Buniello, A., Cerezo, M., Gil, L., Groza, T., Güneş, O., Hall, P., Hayhurst, J., et al. (2023). The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. Nucleic Acids Res. *51*, D977–D985.
3. Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. Science *337*, 1190–1195.
4. Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ward, L.D., et al. (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317–330.
5. Gamazon, E.R., Wheeler, H.E., Shah, K.P., Mozaffari, S.V., Aquino-Michaels, K., Carroll, R.J., Eyler, A.E., Denny, J.C., Nicolae, D.L., Cox, N.J., et al. (2015). A gene-based association method for mapping traits using reference transcriptome data. Nat. Genet. *47*, 1091–1098.
6. Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B.W.J.H., Jansen, R., de Geus, E.J.C., Boomsma, D.I., Wright, F.A., et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. Nat. Genet. *48*, 245–252.
7. Hormozdiari, F., van de Bunt, M., Segrè, A.V., Li, X., Joo, J.W.J., Bilow, M., Sul, J.H., Sankararaman, S., Pasaniuc, B., and Eskin, E. (2016). Colocalization of GWAS and eQTL Signals Detects Target Genes. Am. J. Hum. Genet. *99*, 1245–1260.
8. Meuleman, W., Muratov, A., Rynes, E., Halow, J., Lee, K., Bates, D., Diegel, M., Dunn, D., Neri, F., Teodosiadis, A., et al. (2020). Index and biological spectrum of human DNase I hypersensitive sites. Nature *584*, 244–251.
9. Stingele, S., Stoehr, G., Peplowska, K., Cox, J., Mann, M., and Storchova, Z. (2012). Global analysis of genome, transcriptome and proteome reveals the response to aneuploidy in human cells. Mol. Syst. Biol. *8*, 608.
10. Hassold, T., and Hunt, P. (2001). To err (meiotically) is human: the genesis of human aneuploidy. Nat. Rev. Genet. *2*, 280–291.
11. Lehman, C.D., Nyberg, D.A., Winter, T.C., Kapur, R.P., Resta, R.G., and Luthy, D.A. (1995). Trisomy 13 syndrome: prenatal US findings in a review of 33 cases. Radiology *194*, 217–222.
12. Cereda, A., and Carey, J.C. (2012). The trisomy 18 syndrome. Orphanet J. Rare Dis. *7*, 81.
13. Bull, M.J. (2020). Down Syndrome. N. Engl. J. Med. *382*, 2344–2352.
14. Shao, X., Lv, N., Liao, J., Long, J., Xue, R., Ai, N., Xu, D., and Fan, X. (2019). Copy number variation is highly correlated with differential gene expression: a pan-cancer study. BMC Med. Genet. *20*, 175.
15. Sudmant, P.H., Mallick, S., Nelson, B.J., Hormozdiari, F., Krumm, N., Huddleston, J., Coe, B.P., Baker, C., Nordenfelt, S., Bamshad, M., et al. (2015). Global diversity, population stratification, and selection of human copy-number variation. Science *349*, aab3761.
16. Abel, H.J., Larson, D.E., Regier, A.A., Chiang, C., Das, I., Kanchi, K.L., Layer, R.M., Neale, B.M., Salerno, W.J., Reeves, C., et al. (2020). Mapping and characterization of structural variation in 17,795 human genomes. Nature *583*, 83–89.
17. Collins, R.L., Brand, H., Karczewski, K.J., Zhao, X., Alföldi, J., Francioli, L.C., Khera, A.V., Lowther, C., Gauthier, L.D., Wang, H., et al. (2020). A structural variation reference for medical and population genetics. Nature *581*, 444–451.
18. Almarri, M.A., Bergström, A., Prado-Martinez, J., Yang, F., Fu, B., Dunham, A.S., Chen, Y., Hurles, M.E., Tyler-Smith, C., and Xue, Y. (2020). Population Structure, Stratification, and Introgression of Human Structural Variation. Cell *182*, 189–199.e15.
19. Dauber, A., Yu, Y., Turchin, M.C., Chiang, C.W., Meng, Y.A., Demerath, E.W., Patel, S.R., Rich, S.S., Rotter, J.I., Schreiner, P.J., et al. (2011). Genome-wide Association of Copy-Number Variation Reveals an Association between Short Stature and the Presence of Low-Frequency Genomic Deletions. Am. J. Hum. Genet. *89*, 751–759.
20. Wheeler, E., Huang, N., Bochukova, E.G., Keogh, J.M., Lindsay, S., Garg, S., Henning, E., Blackburn, H., Loos, R.J.F., Wareham, N.J., et al. (2013). Genome-wide SNP and CNV analysis identifies common and low-frequency variants associated with severe early-onset obesity. Nat. Genet. *45*, 513–517.
21. Macé, A., Tuke, M.A., Deelen, P., Kristiansson, K., Mattsson, H., Nõukas, M., Sapkota, Y., Schick, U., Porcu, E., Rüeger, S., et al. (2017).

CNV-association meta-analysis in 191,161 European adults reveals new loci associated with anthropometric traits. Nat. Commun. *8*, 744.

22. Marshall, C.R., Howrigan, D.P., Merico, D., Thiruvahindrapuram, B., Wu, W., Greer, D.S., Antaki, D., Shetty, A., Holmans, P.A., Pinto, D., et al. (2017). Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. Nat. Genet. *49*, 27–35.
23. Aguirre, M., Rivas, M.A., and Priest, J. (2019). Phenome-wide Burden of Copy-Number Variation in the UK Biobank. Am. J. Hum. Genet. *105*, 373–383.
24. Auwerx, C., Lepamets, M., Sadler, M.C., Patxot, M., Stojanov, M., Baud, D., Mägi, R., Estonian Biobank Research Team; Porcu, E., Reymond, A., and Kutalik, Z. (2022). The individual and global impact of copy-number variants on complex human traits. Am. J. Hum. Genet. *109*, 647–668.
25. Hujoel, M.L.A., Sherman, M.A., Barton, A.R., Mukamel, R.E., Sankaran, V.G., Terao, C., and Loh, P.-R. (2022). Influences of rare copy-number variation on human complex traits. Cell *185*, 4233–4248.e27.
26. Collins, R.L., Glessner, J.T., 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.
27. Auwerx, C., Jõeloo, M., Sadler, M.C., Tesio, N., Ojavee, S., Clark, C.J., Mägi, R., Reymond, A., and Kutalik, Z. (2024). Rare copy-number variants as modulators of common disease susceptibility. Genome Med. *16*, 5.
28. Ganna, A., Satterstrom, F.K., Zekavat, S.M., Das, I., Kurki, M.I., Churchhouse, C., Alfoldi, J., Martin, A.R., Havulinna, A.S., Byrnes, A., et al. (2018). Quantifying the Impact of Rare and Ultra-rare Coding Variation across the Phenotypic Spectrum. Am. J. Hum. Genet. *102*, 1204–1211.
29. Chen, C.-Y., Tian, R., Ge, T., Lam, M., Sanchez-Andrade, G., Singh, T., Urpa, L., Liu, J.Z., Sanderson, M., Rowley, C., et al. (2023). The impact of rare protein coding genetic variation on adult cognitive function. Nat. Genet. *55*, 927–938.
30. Huang, Y., Bodnar, D., Chen, C.-Y., Sanchez-Andrade, G., Sanderson, M., Shi, J., Meilleur, K.G., Hurles, M.E., Gerety, S.S., and Tsai, E.A. (2023). Rare genetic variants impact muscle strength. Nat. Commun. *14*, 3449.
31. Backman, J.D., Li, A.H., Marcketta, A., Sun, D., Mbatchou, J., Kessler, M.D., Benner, C., Liu, D., Locke, A.E., Balasubramanian, S., et al. (2021). Exome sequencing and analysis of 454,787 UK Biobank participants. Nature *599*, 628–634.
32. The UK Biobank Whole-Genome Sequencing Consortium (2025). Whole-genome sequencing of 490,640 UK Biobank participants. Nature *645*, 692.
33. Mbatchou, J., Barnard, L., Backman, J., Marcketta, A., Kosmicki, J.A., Ziyatdinov, A., Benner, C., O'Dushlaine, C., Barber, M., Boutkov, B., et al. (2021). Computationally efficient whole-genome regression for quantitative and binary traits. Nat. Genet. *53*, 1097–1103.
34. Brown, M.S., and Goldstein, J.L. (2006). Lowering LDL–Not Only How Low, But How Long? Science *311*, 1721–1723.
35. Goldstein, J.L., and Brown, M.S. (2009). The LDL Receptor. Arterioscler. Thromb. Vasc. Biol. *29*, 431–438.
36. Ogata, T., Matsuo, N., and Nishimura, G. (2001). SHOX haploinsufficiency and overdosage: impact of gonadal function status. J. Med. Genet. *38*, 1–6.
37. Berg, J.J., Li, X., Riall, K., Hayward, L.K., and Sella, G. (2025). Mutation–selection–drift balance models of complex diseases. Genetics *231*, iyaf220.
38. Plenge, R.M., Scolnick, E.M., and Altshuler, D. (2013). Validating therapeutic targets through human genetics. Nat. Rev. Drug Discov. *12*, 581–594.
39. Keren, L., Hausser, J., Lotan-Pompan, M., Vainberg Slutskin, I., Alisar, H., Kaminski, S., Weinberger, A., Alon, U., Milo, R., and Segal, E. (2016). Massively Parallel Interrogation of the Effects of Gene Expression Levels on Fitness. Cell *166*, 1282–1294.e18.
40. Naqvi, S., Kim, S., Hoskens, H., Matthews, H.S., Spritz, R.A., Klein, O.D., Hallgrímsson, B., Swigut, T., Claes, P., Pritchard, J.K., and Wysocka, J. (2023). Precise modulation of transcription factor levels identifies features underlying dosage sensitivity. Nat. Genet. *55*, 841–851.
41. Domingo, J., Minaeva, M., Morris, J.A., Ghatan, S., Ziosi, M., Sanjana, N.E., and Lappalainen, T. (2026). Non-linear transcriptional responses to gradual modulation of transcription factor dosage. eLife *13*.
42. Nadig, A., Replogle, J.M., Pogson, A.N., Murthy, M., McCarroll, S.A., Weissman, J.S., Robinson, E.B., and O'Connor, L.J. (2025). Transcriptome-wide analysis of differential expression in perturbation atlases. Nat. Genet. *57*, 1228–1237.
43. Jacquemont, S., Reymond, A., Zufferey, F., Harewood, L., Walters, R.G., Kutalik, Z., Martinet, D., Shen, Y., Valsesia, A., Beckmann, N.D., et al. (2011). Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus. Nature *478*, 97–102.
44. Golzio, C., Willer, J., Talkowski, M.E., Oh, E.C., Taniguchi, Y., Jacquemont, S., Reymond, A., Sun, M., Sawa, A., Gusella, J.F., et al. (2012). KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. Nature *485*, 363–367.
45. Storey, J.D. (2002). A Direct Approach to False Discovery Rates. J. R. Stat. Soc. Series B Stat. Methodol. *64*, 479–498.
46. Weiss, L.A., Shen, Y., Korn, J.M., Arking, D.E., Miller, D.T., Fossdal, R., Saemundsen, E., Stefansson, H., Ferreira, M.A.R., Green, T., et al. (2008). Association between Microdeletion and Microduplication at 16p11.2 and Autism. N. Engl. J. Med. *358*, 667–675.
47. Golzio, C., and Katsanis, N. (2013). Genetic architecture of reciprocal CNVs. Curr. Opin. Genet. Dev. *23*, 240–248.
48. Auwerx, C., Moix, S., Kutalik, Z., and Reymond, A. (2024). Disentangling mechanisms behind the pleiotropic effects of proximal 16p11.2 BP4-5 CNVs. Am. J. Hum. Genet. *0*.
49. Vitart, V., Rudan, I., Hayward, C., Gray, N.K., Floyd, J., Palmer, C.N.A., Knott, S.A., Kolcic, I., Polasek, O., Graessler, J., et al. (2008). SLC2A9 is a newly identified urate transporter influencing serum urate concentration, urate excretion and gout. Nat. Genet. *40*, 437–442.
50. Anzai, N., Ichida, K., Jutabha, P., Kimura, T., Babu, E., Jin, C.J., Srivastava, S., Kitamura, K., Hisatome, I., Endou, H., and Sakurai, H. (2008). Plasma urate level is directly regulated by a voltage-driven urate efflux transporter URATv1 (SLC2A9) in humans. J. Biol. Chem. *283*, 26834–26838.
51. Mount, D.B., Kwon, C.Y., and Zandi-Nejad, K. (2006). Renal Urate Transport. Rheum. Dis. Clin. North Am. *32*, 313–331.
52. Kaplan, S., Bren, A., Dekel, E., and Alon, U. (2008). The incoherent feed-forward loop can generate non-monotonic input functions for genes. Mol. Syst. Biol. *4*, 203.
53. Kugelman, A., Choy, H.A., Liu, R., Shi, M.M., Gozal, E., and Forman, H.J. (1994). gamma-Glutamyl transpeptidase is increased by oxidative stress in rat alveolar L2 epithelial cells. Am. J. Respir. Cell Mol. Biol. *11*, 586–592.
54. Lee, D.-H., Blomhoff, R., and Jacobs, D.R. (2004). Is serum gamma glutamyltransferase a marker of oxidative stress? Free Radic. Res. *38*, 535–539.
55. Bai, C., Zhang, M., Zhang, Y., He, Y., Dou, H., Wang, Z., Wang, Z., Li, Z., and Zhang, L. (2022). Gamma-Glutamyltransferase Activity (GGT) Is a Long-Sought Biomarker of Redox Status in Blood Circulation: A Retrospective Clinical Study of 44 Types of Human Diseases. Oxid. Med. Cell. Longev. *2022*, 8494076.
56. van Kuilenburg, A.B.P., Meinsma, R., Beke, E., Assmann, B., Ribes, A., Lorente, I., Busch, R., Mayatepek, E., Abeling, N.G.G.M., van Cruchten, A., et al. (2004). beta-Ureidopropionase deficiency: an inborn error of pyrimidine degradation associated with neurological abnormalities. Hum. Mol. Genet. *13*, 2793–2801.

57. Shetewy, A., Shimada-Takaura, K., Warner, D., Jong, C.J., Mehdi, A.-B.A., Alexeyev, M., Takahashi, K., and Schaffer, S.W. (2016). Mitochondrial defects associated with β-alanine toxicity: relevance to hyper-beta-alaninemia. Mol. Cell. Biochem. *416*, 11–22.

58. Dobritzsch, D., Meijer, J., Meinsma, R., Maurer, D., Monavari, A.A., Gummesson, A., Reims, A., Cayuela, J.A., Kuklina, N., Benoist, J.-F., et al. (2022). β-Ureidopropionase deficiency due to novel and rare UPB1 mutations affecting pre-mRNA splicing and protein structural integrity and catalytic activity. Mol. Genet. Metab. *136*, 177–185.

59. 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.

60. Simons, Y.B., Bullaughey, K., Hudson, R.R., and Sella, G. (2018). A population genetic interpretation of GWAS findings for human quantitative traits. PLoS Biol. *16*, e2002985.

61. Sella, G., and Barton, N.H. (2019). Thinking About the Evolution of Complex Traits in the Era of Genome-Wide Association Studies. Annu. Rev. Genomics Hum. Genet. *20*, 461–493.

62. Koch, E., Connally, N.J., Baya, N., Reeve, M.P., Daly, M., Neale, B., Lander, E.S., Bloemendal, A., and Sunyaev, S. (2024). Genetic association data are broadly consistent with stabilizing selection shaping human common diseases and traits. Preprint at bioRxiv. https://doi.org/10.1101/2024.06.19.599789.

63. Patel, R.A., Weiß, C.L., Zhu, H., Mostafavi, H., Simons, Y.B., Spence, J.P., and Pritchard, J.K. (2025). Characterizing selection on complex traits through conditional frequency spectra. Genetics *229*, iyae210.

64. Simons, Y.B., Mostafavi, H., Zhu, H., Smith, C.J., Pritchard, J.K., and Sella, G. (2025). Simple scaling laws control the genetic architectures of human complex traits. PLoS Biol. *23*, e3003402.

65. López-Otín, C., Blasco, M.A., Partridge, L., Serrano, M., and Kroemer, G. (2013). The Hallmarks of Aging. Cell *153*, 1194–1217.

66. Chan, M.S., Arnold, M., Offer, A., Hammami, I., Mafham, M., Armitage, J., Perera, R., and Parish, S. (2021). A Biomarker-based Biological Age in UK Biobank: Composition and Prediction of Mortality and Hospital Admissions. J. Gerontol. A Biol. Sci. Med. Sci. *76*, 1295–1302.

67. Libert, S., Chekholko, A., and Kenyon, C. (2025). A mathematical model that predicts human biological age from physiological traits identifies environmental and genetic factors that influence aging. eLife *13*, RP92092.

68. Schmidt, R., Steinhart, Z., Layeghi, M., Freimer, J.W., Bueno, R., Nguyen, V.Q., Blaeschke, F., Ye, C.J., and Marson, A. (2022). CRISPR activation and interference screens decode stimulation responses in primary human T cells. Science *375*, eabj4008.

69. Hutchinson, A., Asimit, J., and Wallace, C. (2020). Fine-mapping genetic associations. Hum. Mol. Genet. *29*, R81–R88.

70. Lappalainen, T., and MacArthur, D.G. (2021). From variant to function in human disease genetics. Science *373*, 1464–1468.

71. Nasser, J., Bergman, D.T., Fulco, C.P., Guckelberger, P., Doughty, B.R., Patwardhan, T.A., Jones, T.R., Nguyen, T.H., Ulirsch, J.C., Lekschas, F., et al. (2021). Genome-wide enhancer maps link risk variants to disease genes. Nature *593*, 238–243.

72. Connally, N.J., Nazeen, S., Lee, D., Shi, H., Stamatoyannopoulos, J., Chun, S., Cotsapas, C., Cassa, C.A., and Sunyaev, S.R. (2022). The missing link between genetic association and regulatory function. eLife *11*, e74970.

73. Gazal, S., Weissbrod, O., Hormozdiari, F., Dey, K.K., Nasser, J., Jagadeesh, K.A., Weiner, D.J., Shi, H., Fulco, C.P., O'Connor, L.J., et al. (2022). Combining SNP-to-gene linking strategies to identify disease genes and assess disease omnigenicity. Nat. Genet. *54*, 827–836.

74. Gschwind, A.R., Mualim, K.S., Karbalayghareh, A., Sheth, M.U., Dey, K.K., Jagoda, E., Nurtdinov, R.N., Xi, W., Tan, A.S., Jones, H., et al. (2023). An encyclopedia of enhancer-gene regulatory interactions in the human genome. Preprint at bioRxiv. https://doi.org/10.1101/2023.11.09.563812.

75. Mancuso, N., Freund, M.K., Johnson, R., Shi, H., Kichaev, G., Gusev, A., and Pasaniuc, B. (2019). Probabilistic fine-mapping of transcriptome-wide association studies. Nat. Genet. *51*, 675–682.

76. Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M.R., Powell, J.E., Montgomery, G.W., Goddard, M.E., Wray, N.R., Visscher, P.M., and Yang, J. (2016). Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat. Genet. *48*, 481–487.

77. Mostafavi, H., Spence, J.P., Naqvi, S., and Pritchard, J.K. (2023). Systematic differences in discovery of genetic effects on gene expression and complex traits. Nat. Genet. *55*, 1866–1875.

78. Zeng, T., Spence, J.P., Mostafavi, H., and Pritchard, J.K. (2024). Bayesian estimation of gene constraint from an evolutionary model with gene features. Nat. Genet. *56*, 1632–1643.

79. Kelley, D.R., Reshef, Y.A., Bileschi, M., Belanger, D., McLean, C.Y., and Snoek, J. (2018). Sequential regulatory activity prediction across chromosomes with convolutional neural networks. Genome Res. *28*, 739–750.

80. Agarwal, V., and Shendure, J. (2020). Predicting mRNA Abundance Directly from Genomic Sequence Using Deep Convolutional Neural Networks. Cell Rep. *31*, 107663.

81. Avsec, Ž., Agarwal, V., Visentin, D., Ledsam, J.R., Grabska-Barwinska, A., Taylor, K.R., Assael, Y., Jumper, J., Kohli, P., and Kelley, D.R. (2021). Effective gene expression prediction from sequence by integrating long-range interactions. Nat. Methods *18*, 1196–1203.

82. Drusinsky, S., Whalen, S., and Pollard, K.S. (2026). Deep-learning prediction of gene expression from personal genomes. Genome Biol. *27*, 19.

83. Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C.P., Jerby-Arnon, L., Marjanovic, N.D., Dionne, D., Burks, T., Raychowdhury, R., et al. (2016). Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. Cell *167*, 1853–1866.e17.

84. Tewhey, R., Kotliar, D., Park, D.S., Liu, B., Winnicki, S., Reilly, S.K., Andersen, K.G., Mikkelsen, T.S., Lander, E.S., Schaffner, S.F., and Sabeti, P.C. (2016). Direct Identification of Hundreds of Expression-Modulating Variants using a Multiplexed Reporter Assay. Cell *165*, 1519–1529.

85. Kircher, M., Xiong, C., Martin, B., Schubach, M., Inoue, F., Bell, R.J.A., Costello, J.F., Shendure, J., and Ahituv, N. (2019). Saturation mutagenesis of twenty disease-associated regulatory elements at single base-pair resolution. Nat. Commun. *10*, 3583.

86. Cheng, J., Novati, G., Pan, J., Bycroft, C., Žemgulytė, A., Applebaum, T., Pritzel, A., Wong, L.H., Zielinski, M., Sargeant, T., et al. (2023). Accurate proteome-wide missense variant effect prediction with AlphaMissense. Science *381*, eadg7492.

87. Fowler, D.M., and Fields, S. (2014). Deep mutational scanning: a new style of protein science. Nat. Methods *11*, 801–807.

88. GTEx Consortium (2017). Genetic effects on gene expression across human tissues. Nature *550*, 204–213.

89. Jost, M., Santos, D.A., Saunders, R.A., Horlbeck, M.A., Hawkins, J.S., Scaria, S.M., Norman, T.M., Hussmann, J.A., Liem, C.R., Gross, C.A., and Weissman, J.S. (2020). Titrating gene expression using libraries of systematically attenuated CRISPR guide RNAs. Nat. Biotechnol. *38*, 355–364.

90. Joung, J., Ma, S., Tay, T., Geiger-Schuller, K.R., Kirchgatterer, P.C., Verdine, V.K., Guo, B., Arias-Garcia, M.A., Allen, W.E., Singh, A., et al. (2023). A transcription factor atlas of directed differentiation. Cell *186*, 209–229.e26.

91. Chandrasekaran, S.N., Cimini, B.A., Goodale, A., Miller, L., Kost-Alimova, M., Jamali, N., Doench, J.G., Fritchman, B., Skepner, A., Melanson, M., et al. (2024). Three million images and morphological profiles of cells treated with matched chemical and genetic perturbations. Nat. Methods *21*, 1114–1121.

92. Song, B., Liu, D., Dai, W., McMyn, N.F., Wang, Q., Yang, D., Krejci, A., Vasilyev, A., Untermoser, N., Loregger, A., et al. (2025). Decoding heterogeneous single-cell perturbation responses. Nat. Cell Biol. *27*, 493–504.

93. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience *4*, s13742.

94. Abraham, G., Qiu, Y., and Inouye, M. (2017). FlashPCA2: principal component analysis of Biobank-scale genotype datasets. Bioinformatics *33*, 2776–2778.

95. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. Genome Biol. *17*, 122.

96. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature *581*, 434–443.

97. Zhao, M., Wang, Q., Wang, Q., Jia, P., and Zhao, Z. (2013). Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. BMC Bioinf. *14*, S1.

98. Zhu, X., and Stephens, M. (2017). Bayesian large-scale multiple regression with summary statistics from genome-wide association studies. Ann. Appl. Stat. *11*, 1561–1592.

99. Julkunen, H., Cichońska, A., Tiainen, M., Koskela, H., Nybo, K., Mäkelä, V., Nokso-Koivisto, J., Kristiansson, K., Perola, M., Salomaa, V., et al. (2023). Atlas of plasma NMR biomarkers for health and disease in 118,461 individuals from the UK Biobank. Nat. Commun. *14*, 604.

100. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. Nature *562*, 203–209.

101. Behera, S., Catreux, S., Rossi, M., Truong, S., Huang, Z., Ruehle, M., Visvanath, A., Parnaby, G., Roddey, C., Onuchic, V., et al. (2024). Comprehensive genome analysis and variant detection at scale using DRAGEN. Nat. Biotechnol. *43*, 1177–1191.

102. Frankish, A., Carbonell-Sala, S., Diekhans, M., Jungreis, I., Loveland, J.E., Mudge, J.M., Sisu, C., Wright, J.C., Arnan, C., Barnes, I., et al. (2023). GENCODE: reference annotation for the human and mouse genomes in 2023. Nucleic Acids Res. *51*, D942–D949.

103. Conneely, K.N., and Boehnke, M. (2007). So Many Correlated Tests, So Little Time! Rapid Adjustment of *P* Values for Multiple Correlated Tests. Am. J. Hum. Genet. *81*, 1158–1168.

104. Bulik-Sullivan, B.K., Loh, P.-R., Finucane, H.K., Ripke, S., Yang, J., Patterson, N., Daly, M.J., Price, A.L., and Neale, B.M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. *47*, 291–295.

105. Delyon, B., Lavielle, M., and Moulines, E. (1999). Convergence of a Stochastic Approximation Version of the EM Algorithm. Ann. Statist. *27*, 94–128.

106. Kuhn, E., and Lavielle, M. (2004). Coupling a stochastic approximation version of EM with an MCMC procedure. ESAIM P. S. *8*, 115–131.

107. Spence, J.P., Sinnott-Armstrong, N., Assimes, T.L., and Pritchard, J.K. (2022). A flexible modeling and inference framework for estimating variant effect sizes from GWAS summary statistics. Preprint at bioRxiv. https://doi.org/10.1101/2022.04.18.488696.

108. Morgante, F., Carbonetto, P., Wang, G., Zou, Y., Sarkar, A., and Stephens, M. (2023). A flexible empirical Bayes approach to multivariate multiple regression, and its improved accuracy in predicting multi-tissue gene expression from genotypes. PLoS Genet. *19*, e1010539.

109. Kunkel, D., Sørensen, P., Shankar, V., and Morgante, F. (2025). Improving polygenic prediction from summary data by learning patterns of effect sharing across multiple phenotypes. PLoS Genet. *21*, e1011519.

110. Stephens, M. (2016). False discovery rates: a new deal. Biostat, kxw041.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Gene-level burden tests	This study	https://zenodo.org/records/19140003
Trait-level summary statistics	This study	https://zenodo.org/records/19140003
Trait measurements	UK Biobank	https://www.ukbiobank.ac.uk/
Protein quantification	UK Biobank	https://www.ukbiobank.ac.uk/
NMR metabolite quantification	UK Biobank	https://www.ukbiobank.ac.uk/
Whole-exome sequencing and whole-genome sequencing	UK Biobank	https://www.ukbiobank.ac.uk/
GWAS summary statistics	Neale lab	https://www.nealelab.is/uk-biobank
GENCODE v39	GENCODE	https://www.gencodegenes.org/human/release_39.html
Estimates of gene-level constraint and misannotation probability	Zeng et al.[78]	https://zenodo.org/records/10403680
Software and algorithms		
Data processing	This study	https://zenodo.org/records/19140111
Monotonicity inference	This study	https://zenodo.org/records/19140111
Trait buffering inference	This study	https://zenodo.org/records/19140111
PLINK 2.0	Chang et al.[93]	https://www.cog-genomics.org/plink/2.0/
FlashPCA2	Abraham et al.[94]	https://github.com/gabraham/flashpca
Ensembl Variant Effect Predictor (VEP)	McLaren et al.[95]	https://github.com/Ensembl/ensembl-vep
LOFTEE	Karczewski et al.[96]	https://github.com/konradjk/loftee
REGENIE	Mbatchou et al.[33]	https://github.com/rgcgithub/regenie
Python	Python Software Foundation	https://www.python.org/
R	R Foundation	https://www.r-project.org/

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Inclusion and ethics

This study relied on data provided by participants in the United Kingdom to the UKB. The UKB is a large biomedical database that provides non-exclusive access to individual-level data. Primary administration of the UKB is handled by groups within the United Kingdom, and a local ethics board provides oversight over the collection of data. The UKB has approval from the North West Multi-centre Research Ethics Committee as a Research Tissue Bank (RTB) approval. This approval means that researchers do not require separate ethical clearance and can operate under the RTB approval.

The main burden tests performed in this paper do not subset individuals based on measures of genetic similarity. However, to compare the effects of confounding and stratification between our study and previous studies, we subset individuals based on self-identification as "White British" or "non-British White'" and genetic principal components (PCs) for a small set of burden tests presented in Methods S1, appendix H.1.1. Results and summary statistics from all analyses presented here are provided as a resource to the scientific community (data and code availability).

Sex was used as a covariate in genome-wide burden tests. The UK Biobank consists of samples from approximately half a million individuals. Traits that were used to perform burden tests have variable sample sizes, which are reported in their summary statistics. No methods were used to pre-determine the sample size. We only analyzed individuals with whole-exome sequencing available in the UK Biobank. In addition, we excluded individuals that did not pass quality control filters for copy-number variant genotyping, which is described in our Methods S1. When aggregating variants into burden tests, we only included potential loss-of-function variants that passed certain quality measures as described in our STAR Methods and Methods S1.

METHOD DETAILS

The goal of our data analysis approach was to estimate burden summary statistics for LoF variants, whole-gene deletions, whole-gene duplications, and partial deletions using a consistent set of decisions and tools. This consistency between the different burden tests allowed us to compare effect sizes across the variant classes. We additionally performed burden tests using synonymous variants as a negative control to estimate any residual confounding (Methods S1, appendix H.1). We improved upon previous LoF burden approaches by using variable frequency filters based on gene constraint (Methods S1, appendix H.1.3). We also improved upon prior duplication burden tests in the UKB[23–25,27] by using copy number calls from WGS data rather than microarray data. CNV calls from WGS are more accurate, have higher resolution of breakpoints, and can better detect rare CNVs.[97]

To make our approach amenable to researchers without access to individual level data, we modeled the summary statistics rather than the underlying genotype data. We used the regression with summary statistics (RSS) likelihood,[98] which provides a likelihood model for association test summary statistics. To our knowledge, this is the first use of the RSS likelihood for burden regression. To demonstrate its applicability and correctness, we performed extensive simulations, which are cataloged in Methods S1, appendix G.

Phenotypes

We curated 410 continuous phenotypes in the UKB on which to perform burden tests. We filtered the phenotypes used for the main analysis in the paper as follows. We required phenotypes to have measurements in at least 40K individuals. We excluded any traits where more than 50% of the instances match the mode of the trait to remove quasi-categorical traits, which is more relaxed compared to Backman et al.,[31] who excluded traits with a cutoff of 20%. To perform inference, we chose traits where at least 7,500 genes had gene-level burden test estimates for both LoF variants and duplications. Finally, we required that either the mean squared effect size of the LoF burden tests or the duplications was nominally significant at the $\alpha = 0.05$ level for each trait. This removes traits where there is not enough signal to perform inference. After filtering, 94 traits remained and were used throughout the main analysis of the paper.

For phenotypes that were collected more than once in returning visits, we used the first instance of the observation for each individual. Backman et al.[31] took the mean across such instances, but we believe that this would reduce the noise for some individuals more than others in a not-completely-at-random fashion. It would also make it challenging to interpret the relationship between mean phenotype and covariates that are observed at specific instances, such as age. Some of our traits had arrayed measurements, but they were all taken in the same visit so we decided to use the mean value across the arrayed items.

Separately, we acquired access to the full 500K baseline metabolomics data from the nuclear magnetic resonance (NMR) metabolomics dataset generated by Nightingale Health PLC.[99]

Genotypes

We included 15 genotyping PCs, estimated previously in the UKB.[100] Backman et al.[31] performed ancestry-specific analyses and included 10 genotyping PCs for each ancestry. Since we used all 460K WES samples, we included an additional 5 PCs to address potential stratification. Our analysis of synonymous variants (Methods S1, appendix H.1.1) supported this choice.

Following the lead of Backman et al.,[31] we included 20 rare variant PCs, by which we mean PCs derived from rare WES variants. We subsampled 300K variants uniformly at random from the rare fraction, which we defined as variants with minor allele count (MAC) greater than 20 and minor allele frequency (MAF) less than 1%. We used an approximate principal components analysis (PCA) algorithm implemented in FlashPCA2.[94]

We restricted our burden analysis to LoF sites with MAF less than 1%. High-confidence LoF sites were annotated using Ensembl's Variant Effect Predictor[95] with the LOFTEE plugin.[96] LOFTEE filters remove annotated LoF sites that might escape nonsense-mediated decay (NMD). Instead of using progressively stricter MAF cutoffs as in Backman et al.,[31] we used all LoF sites with MAF less than 1% but required that the misannotation probability be less than 10%. The misannotation probability[78] takes into account both the frequency of the site and the constraint experienced by the gene to determine a posterior probability of being misannotated as a LoF. These probabilities were previously reported for LoF-introducing single nucleotide polymorphisms (SNPs) for genes with estimated s_{het} values.[78] A small number of genes do not have s_{het} estimates, and not all variants in the WES data were SNPs. For genes with missing s_{het} values, s_{het} was imputed using the mean s_{het} value. Then, for all LoF variants missing a misannotation probability, misannotation probabilities were imputed using k-nearest neighbors (KNN) regression with 10 nearest neighbors as determined by allele frequency and s_{het}. A 10% cutoff on this probability retains 96.46% of variants. We demonstrated that relative to standard frequency cutoffs, using a misannotation probability cutoff increases the signal in the burden tests without increasing the size of standard errors (Methods S1, appendix H.1.3).

CNVs were previously called in the UKB using Illumina's DRAGEN software.[32,101] A deletion was defined as any loss that affected the entire gene body with an additional 1 Kbp flanking region. A duplication was defined as any gain that affected the entire gene body with an additional 1 Kbp flanking region. Any deletions that affected only part of the gene body were categorized as potential loss-of-function (pLoF) alleles. For deletions and pLoF variants, the alternative allele coded the number of copy loss events. For duplications, a copy number of two was coded as the reference homozygote and a copy number greater than two was coded as a heterozygote. While it is theoretically possible that individuals could carry arbitrarily many copies of a gene, in practice we found that individuals coded as heterozygous in this scheme almost always carried three copies (Methods S1, appendix H.3). We observed that some

samples had a large amount of genome-wide copy gain or loss, which is likely a genotyping error (Methods S1, appendix H.4). To account for this, we removed any sample with greater than 10 Mbp of deleted or duplicated genome sequence.

We used GENCODE version 39 to define gene intervals.[102] We excluded genes on the Y chromosome and the mitochondrial genome.

Burden tests

Burden genotypes are composite genotypes consisting of multiple distinct variants that have the same putative downstream dosage effect. For LoF burden tests, we combined different LoF variants in the same gene into one composite genotype (Methods S1, appendix H.2). Similarly, for CNVs, we combined different variants affecting the same gene with the same expected dosage effect (e.g., duplications) into one composite genotype.

We used REGENIE to perform gene-level burden tests.[33] For the whole-genome regression, which is the first step in REGENIE, SNPs from the genotyping array were pruned with a 1000 variant sliding window with 100 variant shifts and an R^2 threshold of 0.9. SNPs were then filtered to have MAF >1%, genotype missingness <10%, and Hardy-Weinberg equilibrium (HWE) test p-value $>10^{-15}$.

For LoF burden tests, we used the following covariates: age, sex, age-by-sex, age squared, 15 genotyping PCs, 20 rare variant PCs, and WES batch. For the duplication, deletion, or pLoF burden tests, we used the following covariates: age, sex, age-by-sex, age squared, 15 genotyping PCs, 20 rare variant PCs, and WGS batch. When analyzing NMR metabolites, we additionally included the spectrometer and processing batch as covariates. We used the default 'max' mask to create burden genotypes.

After whole-genome regression, we used the second step of REGENIE to perform burden tests. Phenotypes were rank-inverse-normal transformed in both the first and second step. The same covariates were used in both steps.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistics and reproducibility

Genome-wide burden tests performed in this study follow quality control standards set by prior analyses, with additional quality control measures discussed in Methods S1 Appendix H. Results from parameter inference are paired with appropriate measures of uncertainty. Inclusion and exclusion criteria during analysis are discussed wherever appropriate. We provide code to replicate our analysis (data and code availability).

Monotonicity model

We modeled the standardized trait of interest $\mathbf{Y}$ in N individuals as $\mathbf{Y} = \mathbf{X}\boldsymbol{\gamma}_1 + \mathbf{Z}\boldsymbol{\gamma}_2 + \epsilon$, where $\mathbf{X}$ and $\mathbf{Z}$ are the burden genotypes from LoF variants and duplications respectively. The error ϵ was assumed to be drawn from a normal distribution. Polygenic signal, population stratification, and other sources of confounding were accounted for when performing the burden tests by including covariates. We built a hierarchical model by specifying a distribution on the latent effect sizes of the jth gene:

$$\begin{bmatrix} \gamma_{1j} \\ \gamma_{2j} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \bar{\gamma}_1 \\ \bar{\gamma}_2 \end{bmatrix}, \begin{bmatrix} \sigma_{11}^2 & \sigma_{12} \\ \sigma_{12} & \sigma_{22}^2 \end{bmatrix} \right).$$

This allowed us to model the quantities of interest. The average effects of deletions and duplications are represented by $\bar{\gamma}_1\mathbf{1}$ and $\bar{\gamma}_2\mathbf{1}$ respectively. We defined the monotonicity, $\phi \in [-1,1]$, to be a quantity that is proportional to the negative uncentered second moment of the latent effect sizes, $\phi \propto -\mathbb{E}\left[\gamma_{1j}\gamma_{2j}\right]$. Thus, if genes with large effects tend to have opposite signs on average, ϕ will be closer to 1. Similarly, if genes with large effects tend to have the same sign on average, ϕ will be closer to -1. The exact definition of the parameter is provided in Methods S1, appendix B.2.

The observed burden effect sizes for the jth gene, $\hat{\gamma}_{1j}$ and $\hat{\gamma}_{2j}$, are noisy estimators of the underlying latent effect size. Furthermore, due to correlation between duplication burden genotypes, $\hat{\gamma}_{2j}$ is not an unbiased estimator of γ_{2j} because the observed effect size is inflated by the effect of nearby genes. To account for these concerns, we used the RSS likelihood,[98] which formalizes an approximation for regression-based summary statistics that is commonly used in statistical genetics.[103,104] An extensive description of the model is provided in Methods S1, appendix B.

We developed unbiased estimators of $\bar{\gamma}_1$ and $\bar{\gamma}_2$. Because ϕ is a complex function of the parameters of the prior, it was challenging to derive unbiased estimators. Instead, we opted to use maximum likelihood estimation for ϕ, which guarantees consistency. Inference and simulations under this model are described in Methods S1, appendices C and G.1 respectively.

Trait buffering model

Since the monotonicity, ϕ, is a function of first and second moments of the latent effect sizes, we used a normal distribution as the prior. However, our buffering model proposed a more complex hypothesis about the distribution of the latent effect sizes, which required a more flexible prior over the latent effect sizes.

To address this, we modified our monotonicity model to use the MASH model prior,[59] which is a flexible multivariate prior. MASH uses a mixture of K multivariate normal distributions to model the latent space:

$$\begin{bmatrix} \gamma_{1j} \\ \gamma_{2j} \end{bmatrix} \sim \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{0}, \mathbf{V}_k) \ .$$

Since the true latent distribution is unknown, MASH uses a grid of fixed covariance matrices and learns the mixture weights $\boldsymbol{\pi}$ using empirical Bayes. In its original implementation, MASH assumes independence between observed samples, allowing for efficient optimization. Because the RSS likelihood necessarily induces dependencies between the observations via correlation between the burden genotypes, we implemented a stochastic approximation expectation maximization (SAEM) algorithm to optimize $\boldsymbol{\pi}$.[105,106] Alternative approaches to this model that use variational inference have been proposed.[107–109]

As we defined it, the amount of trait buffering can be estimated by taking the dot product of the latent effect size with the normal vector of the hyperplane separating the two types of buffering, which we explain further in Methods S1, appendix D.2. We define the measure of trait buffering as

$$\xi = \frac{1}{M} \sum_{j=1}^{M} \log_2 \left(\frac{3}{2}\right) \gamma_{1j} + \gamma_{2j},$$

where M is the number of genes. We perform inference using the posterior distribution for ξ given the observed summary statistics. To determine how much is contributed to ξ by monotone or non-monotone genes, we defined

$$\xi_{\mathrm{m}} = \frac{1}{M} \sum_{j:\gamma_{1j}\gamma_{2j}<0} \left[\log_2 \left(\frac{3}{2}\right) \gamma_{1j} + \gamma_{2j}\right]$$

$$\xi_{\mathrm{nm}} = \frac{1}{M} \sum_{j:\gamma_{1j}\gamma_{2j}>0} \left[\log_2 \left(\frac{3}{2}\right) \gamma_{1j} + \gamma_{2j}\right].$$

The intuition here is that ξ_{m} sums up contributions from monotone genes (where $\gamma_{1j}\gamma_{2j} < 0$), and ξ_{nm} sums up contributions from non-monotone gene (where $\gamma_{1j}\gamma_{2j} > 0$). Note that, by construction, $\xi = \xi_{\mathrm{m}} + \xi_{\mathrm{nm}}$. ξ is zero under the null model where GDRCs are equally likely to be buffered against increasing or decreasing the trait. The details of the model and inference can be found in Appendices D and E respectively.

Data analysis

Throughout, we used unbiased maximum likelihood estimates (MLEs) for the duplication burden estimates when they are displayed in the main text, which account for correlations with neighboring genes.

In Figure 1B, we used total least squares to fit a regression line with no intercept. Weights in the regression corresponded to the inverse variance of the estimates. To provide a better fit, we used multiple initial parameter values. We retrieved GWAS summary statistics from the UKB to generate trumpet plots (Figure 1C and Methods S1, appendix I1). The collection and analysis of these data is described in Methods S1, appendix I.2. We fit a LOESS curve to visualize the conditional mean of the derived allele effect given the derived allele frequency.

For the analysis presented in Figure 5, we use the local false sign rate (LFSR).[110] For a given burden test effect size estimate for a gene, a rate of less than 10% implies that more than 90% of the posterior density is on one side of zero, indicating confidence in the sign of the effect.