medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

Buffering and non-monotonic behavior of gene dosage response curves for human complex traits

Nikhil Milind [1]*, Courtney J. Smith [1], Huisheng Zhu [2], Jeffrey P. Spence [1]*, Jonathan K. Pritchard [1,2]*

Abstract

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, at individual genes, reducing expression should have the opposite effect on a phenotype as increasing expression. To understand this paradox, we introduce a concept called the gene dosage response curve (GDRC) that 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 and, surprisingly, that as many as 40% of GDRCs are non-monotone, with large increases and decreases in expression affecting the trait in the same direction. 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.

[1]Department of Genetics, Stanford University, CA, USA
[2]Department of Biology, Stanford University, CA, USA
*Corresponding authors: nmilind@stanford.edu, jspence@stanford.edu, pritch@stanford.edu

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

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 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 a few 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 to traits [19–22, 24, 25, 27–29]. An individual's genome-wide burden is estimated by counting the total number of CNVs carried by an individual, 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 [20–22]. That is, increasing expression has the same effect as decreasing expression on average (Appendix A). This seemingly contradicts the intuition that, for a given gene, increasing expression should have the opposite effect of 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) [30, 31].

Our analysis of 91 continuous traits revealed that, across traits, the average effects of deletions and duplications are often non-zero and usually affect the trait in the same direction. Although the data confirm the intuition that for most genes deletions have opposite effects to duplications, we identified a substantial number of cases where deletions and duplications of a single gene affect a trait in the same direction. To explain these counterintuitive observations, we introduce the concept of gene dosage response curves (GDRCs), study their properties, and develop a simple model of their evolution.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

Results

Gene-level burden tests for loss-of-function variants and duplications

To better understand how variants with different effects on dosage impact traits, we performed burden tests for 410 continuous traits using LoF variants, deletions, and duplications in the UKB. In the analysis presented, we focused on LoF variant burden tests because they are better powered than deletions, but deletions showed broadly similar patterns of effect (Appendix F). 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 [32]. Continuous traits were selected from various blood biochemistry, blood count, blood metabolite, and anthropometric measurements available in the UKB (Data Availability). We accounted for sex, age, batch effects, and other sources of confounding by including appropriate covariates (Methods) and performed quality control steps (Appendix F). The summary statistics from these burden tests are provided as a resource to the community (Data Availability). We selected a subset of 91 continuous traits for analysis in the paper (Data Availability).

Gene dosage response curves could explain paradoxical patterns in burden tests

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 here as 'gene dosage', are often 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]. However, LoF and duplication burden tests assess phenotypic effects at opposite ends of the dosage spectrum and under a linear model would have opposite directions of effect. Thus, a linear model is incompatible with the genome-wide burden effects of LoF variants and duplications having the same sign (Appendix A). Recent work on the effects of gene dosage on molecular traits shows these relationships are often nonlinear [33, 34]; as we will show below, an assumption of nonlinearity is necessary but not sufficient to explain these observations.

We next extended the observations of directionality by estimating the effect of each individual gene on the trait using burden tests, and then averaging those estimates across all genes for 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). Curiously, however, when we look at top hits from GWASs, we find no evidence for such an effect (Figures 1C and F.8).

Taken together, these observations provide contradictory information about the relationship between dosage and trait. To explain these observations, we need to understand the gene-level heterogeneity present in the dosage-response relationships. We envision the relationship between gene expression and trait as a gene dosage response curve (GDRC): the continuous relationship between gene expression and 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 result in different trait values on average. A given

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

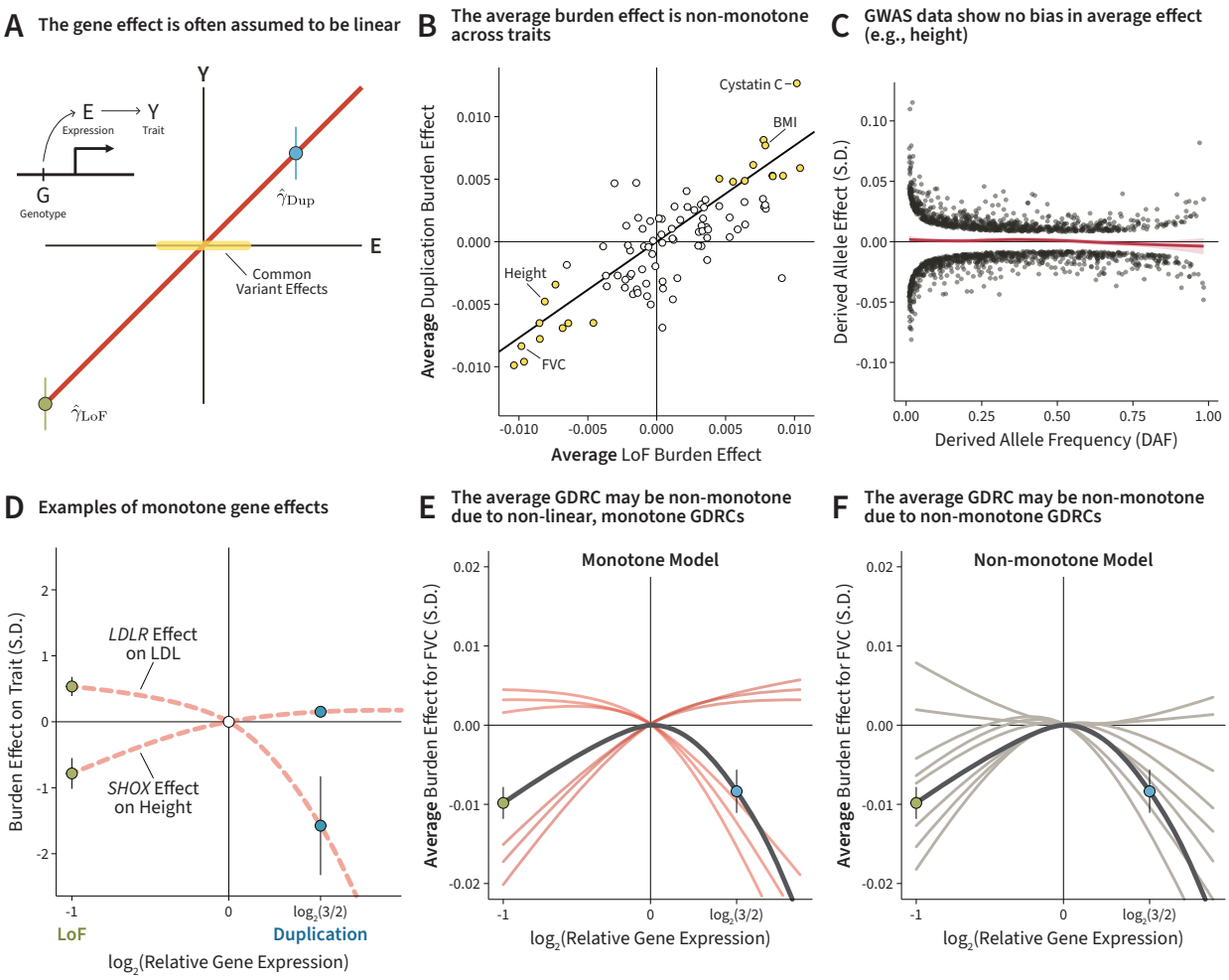

FIGURE 1. Explaining the paradoxical observations in burden data. A. The hypothetical relationship between gene dosage (E) and expected trait value (Y) is often assumed to be linear. LoF variants and duplications estimate the gene effect at different dosage values ($\hat{\gamma}_{\mathrm{LoF}}$ and $\hat{\gamma}_{\mathrm{Dup}}$). **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 ($\beta = 0.77, z = 17.09$). **C.** Conditionally independent GWAS top hits for height. The LOESS curve shows no bias across the frequency spectrum. **D.** Burden estimates and 95% confidence intervals for two gene-trait pairs, with hypothetical GDRCs. **E.** The first model explains the non-monotone average GDRC of FVC (with 95% confidence intervals) using monotone gene curves. **F.** The second model explains the non-monotone average GDRC of FVC (with 95% confidence intervals) using non-monotone gene curves.

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 acts as a conceptual framework to unify various estimates of gene effects, 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 part of the GDRC. We define the origin to be the mean trait and expression value in the population (Appendix A). A burden estimate of the LoF effect, $\hat{\gamma}_{\mathrm{LoF}}$, is an estimate of a point on the left tail of the curve, while a burden estimate of the duplication effect, $\hat{\gamma}_{\mathrm{Dup}}$, is an estimate of a

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

point on the right tail of the curve (Figure 1A). The effect estimated by TWAS, $\hat{\gamma}_{\mathrm{TWAS}}$, is the slope of a linear approximation to the GDRC in the region of expression levels spanned by common expression quantitative trait loci (eQTL). While we visualize LoF variants and duplications as having specific dosage effects on expression (50% and 150% respectively), our theoretical arguments will require only that they generally decrease, or increase, expression respectively to understand the general characteristics of GDRCs.

Intuition suggests that GDRCs should be monotone, meaning that reducing and increasing expression should have opposite phenotypic effects (Figure 1D) [24, 25, 27, 35, 36]. The genome-wide average burden effects that we computed can be interpreted as points on the average GDRC obtained by averaging the GDRC of each gene for a given trait. If average deletion and duplication effects are in the same direction, it would imply a non-monotone GDRC. It therefore seems surprising that aggregating presumably monotone GDRCs would result in a non-monotone average GDRC.

We suggest two models that could explain observations of non-monotone average GDRCs. In the *monotone model*, the non-monotone average GDRC can be explained using only monotone gene-level GDRCs if the monotone curves are systematically buffered against one trait direction (Figure 1E). In the *non-monotone model*, a subset of GDRCs are non-monotone, and are more often non-monotone in a particular direction, driving the average GDRC to be non-monotone (Figure 1F). These models are not mutually exclusive and may both contribute to the non-monotone average GDRC observed for a trait.

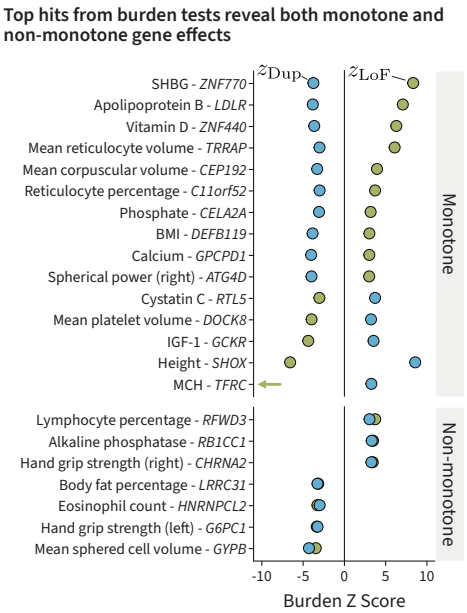

FIGURE 2. Top hits from burden tests. The Z scores from regression-based burden tests using LoF and duplication variants across various continuous traits. Top hits are separated into monotone and non-monotone relationships. The green arrow represents a Z score smaller than -10.

To understand the relative contribution of these models to the average burden effects, we started by looking at top genes for various traits (Figure 2). As we expected, most top genes have monotone effects on traits. However, surprisingly, a sizable minority of the top hits had non-monotone effects. Thus, both

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

models could plausibly contribute to the observed average GDRCs.

Most traits are monotone on average

To better understand the contribution of our proposed models to the non-monotone average burden effects (Figures 1E and 1F), we wanted to understand the overall monotonicity of GDRCs, beyond just significant hits.

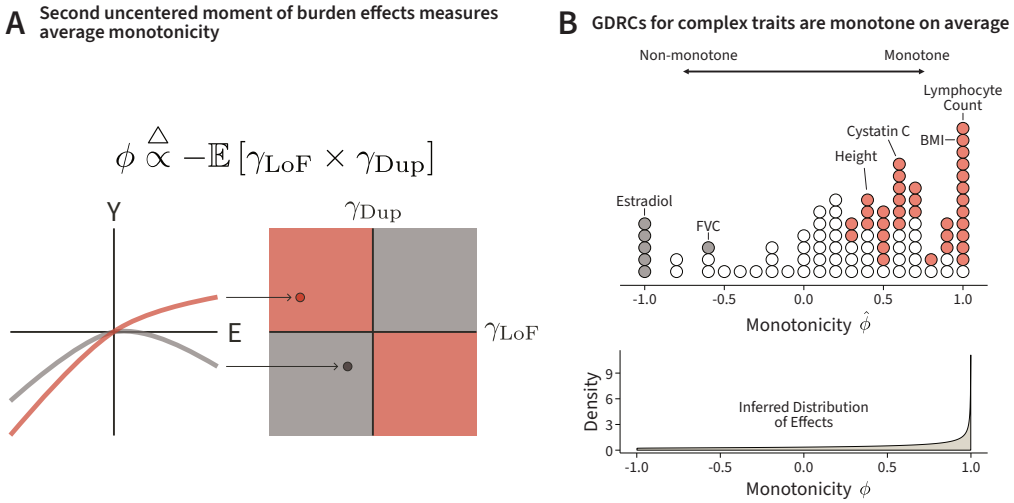

Figure 3. Monotonicity inference for traits. A. Monotone curves (red) and non-monotone curves (gray) fall in different quadrants on the burden effect plot. A natural estimate of comparing points in these quadrants is the second uncentered moment between burden effect sizes. **B.** Estimated values of ϕ. Red and gray points are significantly non-zero ($p < 0.05$). We inferred a distribution over the latent monotonicity values, which has more density over positive monotonicity values.

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 would 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 and term this the *monotonicity* (Methods, Appendix B). ϕ is positive if genes with large effects on a trait tend to have monotone effects on average.

Across traits, we found that the burden signal was explained predominantly by monotone GDRCs, although five traits had significantly negative monotonicity estimates (Figures 3B and F.9). To get a sense of the average level of monotonicity across traits, we inferred a distribution over the monotonicity estimates across traits, which confirmed that both positive and negative ϕ may be observed (Figure 3B). 71% of the density of the inferred distribution was over positive values of ϕ, while the rest was negative.

Our monotonicity estimates do not explain the non-monotone average burden effect. For example, while forced vital capacity (FVC) has a high degree of non-monotonicity and a non-monotone average GDRC, height and cystatin C both have a high degree of monotonicity but also have non-monotone av-

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

erage GDRCs (Figure 3B). However, the predominance of positive values of ϕ suggests that most of the overall burden signal comes from monotone genes. This matches our biological expectations of monotone gene-level effects, and is potentially consistent with the monotone model.

Non-monotone genes align with the direction of average effects

The distribution of monotonicity estimates suggests that the genome-wide signal is composed of both monotone and non-monotone GDRCs. To systematically assess if we could detect non-monotone GDRCs, we ascertained genes with marginally significant LoF and duplication effects ($|z| > 2$) in any trait. We identified 633 gene-trait association pairs across all of our traits, of which 287 (45.3%) exhibited a non-monotone relationship (Figure 4A). This suggests that the burden data is potentially consistent with the non-monotone model.

The non-monotone model (Figure 1F) proposes that there are more non-monotone curves aligned with the average burden effect than not. To assess this, we polarized ascertained gene-trait pairs based on the direction of the average burden effect. Consistent with our model, non-monotone genes were significantly more likely to be found in the quadrant concordant with the average effect (Figures 4A and 4B). As a specific example of a gene that falls in the concordant quadrant, the average LoF and duplication burden effect is to reduce hand grip strength (Table F.1), and one of the top non-monotone genes, *G6PC1*, reduces hand grip strength upon perturbation in either direction (Figure 2).

The presence of non-monotone GDRCs that contribute strongly to the average burden effect is surprising. We can use a simple mechanistic model to see why such non-monotone GDRCs are unexpected. Consider a gene that affects a trait via a single biological pathway (Figure 4C). Intuition about biological mechanisms suggests that each step along a 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 (Figure 4C). Similarly, perturbations of the expression levels of transcription factors can result in non-linear changes in the expression of their targets [33, 34]. As the dosage effect percolates through a pathway towards 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 4C, Appendix A).

However, non-monotone GDRCs can arise if a gene's effects flow through two or more pathways (Figure 4D). 3-hydroxybutyrate levels, which increase in blood during ketogenesis, provide an interpretable example. *CYP11B1* has a non-monotone relationship with 3-hydroxybutyrate levels in the blood, with both LoF variants and duplications increasing 3-hydroxybutyrate on average ($\hat{\gamma}_{\text{LoF}} = 0.15, z_{\text{LoF}} = 3.03, \hat{\gamma}_{\text{Dup}} = 0.68, z_{\text{Dup}} = 2.66$). *CYP11B1* encodes a key adrenal enzyme involved in the production of cortisol. Deficiency of the enzyme encoded by *CYP11B1*, referred to as 11β-hydroxylase deficiency, is a cause of a congenital adrenal condition that results in reduced cortisol levels and subsequent hypoglycemia [37]. Without sufficient glucose levels, the body generates increased amounts of ketones,

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

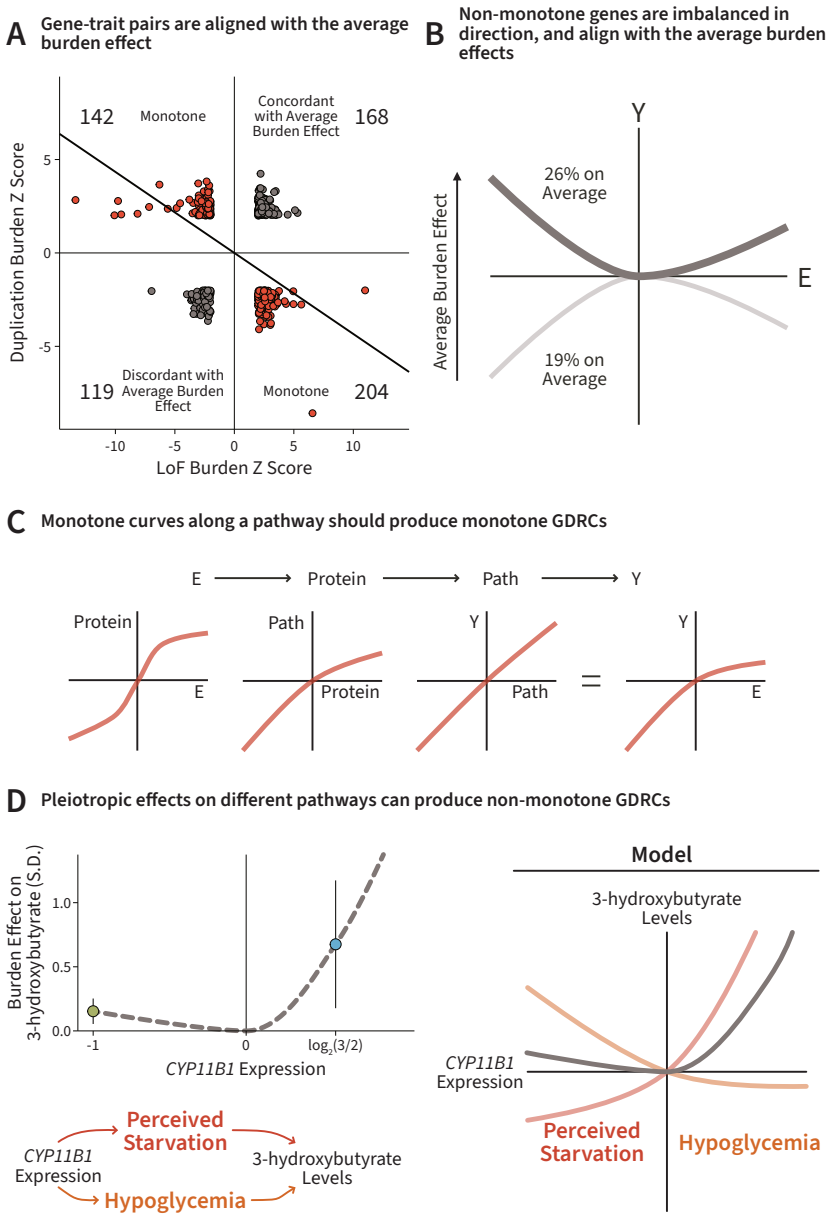

FIGURE 4. Large fraction of non-monotone GDRCs. A. Top gene-trait pairs ($|z| > 2$) for traits with significant average burden effects (Table F.1). The burden effects were polarized by the direction of the average burden effect. **B.** The proportion of concordant versus discordant non-monotone GDRCs per trait show that non-monotone GDRCs tend to align themselves with the average burden effect (Student's t-test $p = 1.59 \times 10^{-2}$). **C.** Monotone curves along a pathway will compose to form a monotone GDRC. **D.** The effect of *CYP11B1* propagates through two separate paths to affect 3-hydroxybutyrate levels, which may explain the non-monotone GDRC.

especially 3-hydroxybutyrate, to serve as an alternative energy source [38]. Meanwhile, duplications in *CYP11B1* can lead to increased production of cortisol, which over time can lead to the development of Cushing's syndrome and insulin resistance [39, 40]. This insulin resistance can then lead to a state of perceived hypoglycemia, resulting in increased ketone levels, especially increased 3-hydroxybutyrate, in the blood [41]. Generalizing these biological principles, non-monotone GDRCs can arise whenever the

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

effect of a gene propagates through multiple pathways to affect the trait (Figure 4D).

Overall, non-monotone GDRCs exist and can have reasonable biological underpinnings. The number of non-monotone GDRCs is imbalanced, and suggestively aligns with the average burden effect across traits. This provides support that our second model contributes to the average burden effect.

Genes are buffered against one trait direction

The analysis of monotonicity and the imbalance in non-monotone GDRCs suggest that both the monotone and non-monotone models are plausible explanations of the non-monotone average GDRCs. If the mixture of GDRCs for the trait are depressed in one direction, then the average GDRC will be non-monotone (Figure 5A). Another interpretation of these curves 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. We refer to this as *trait buffering*. We endeavored to quantify how much each of our models contributed to the non-monotone average GDRC by measuring and decomposing the trait buffering for various traits.

In the burden effect plot, curves experiencing trait buffering preferentially map to a region that is away from a diagonal line (Figure 5A). The diagonal line represents perfectly linear GDRCs (Appendix E). 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 [42], which models the joint distribution of γ_{LoF} and γ_{Dup} as a mixture of bivariate normal distributions. This prior distribution is learned from the data. We used posterior samples of γ_{LoF} and γ_{Dup} for each gene to better understand the distribution of genes on the burden effect plot. As an example, the posterior means of γ_{LoF} and γ_{Dup} for height are displayed in Figure 5B. Both monotone and non-monotone GDRCs are predominantly below the diagonal, concordant with the direction of the average burden effect.

To formalize our intuition, we developed a statistic to measure this trait buffering, ξ, that aggregates the signed strength of deviation from the diagonal line across all GDRCs (Appendix E). We calculate the statistic using posterior samples of γ_{LoF} and γ_{Dup}.

The sign of ξ indicates which direction is being buffered against. For example, our model predicts that height should have a negative value of ξ and cystatin C should have a positive value of ξ. 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 both the average LoF and duplication burden effects (Figure 5C).

To understand the extent to which our proposed models (Figures 1E and 1F) contributed to trait

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

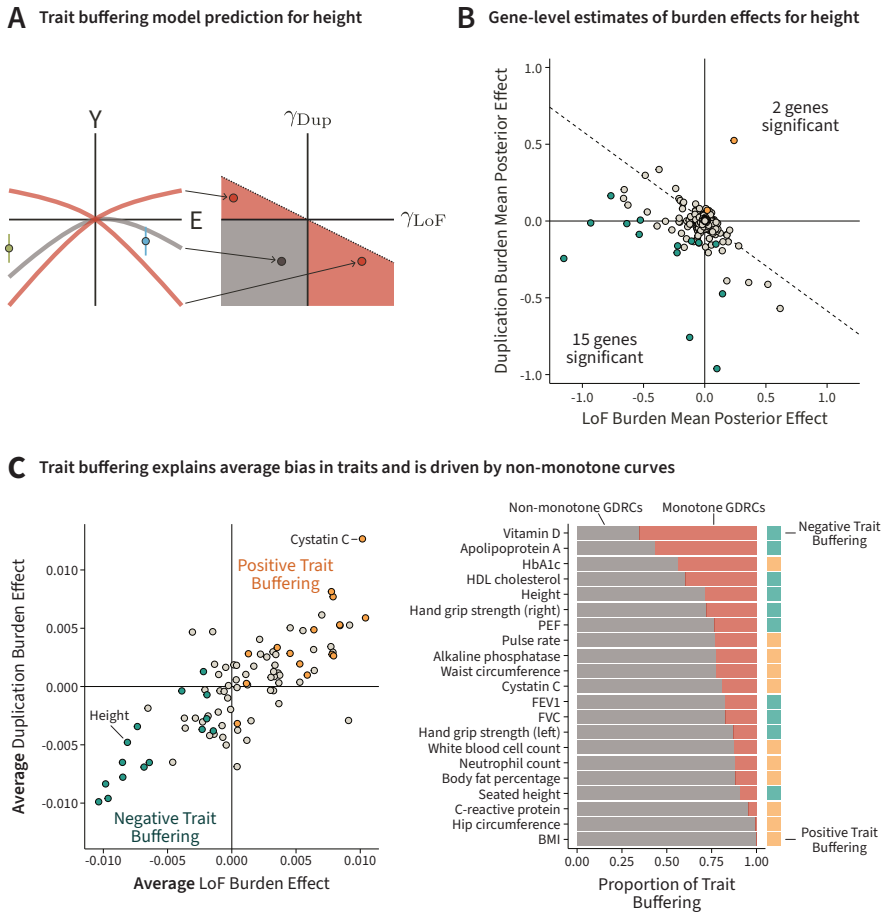

FIGURE 5. Trait buffering as a model for GDRC architecture. A. Combining both models of GDRCs contributing to the average burden effect. The GDRCs fall below a diagonal line for traits with negative average burden effects such as height. The average loss-of-function and duplication burden effects will be in the direction of trait buffering (green and blue points). **B.** Estimates of the latent burden effects show, visually, that height is consistent with trait buffering. Yellow and green genes have a local false sign rate less than 0.05. **C.** (Left) The measure of trait buffering is strongly correlated with the average burden effects. Yellow and green genes have a local false sign rate less than 0.05. (Right) When we decompose the trait buffering measure, both monotone and non-monotone genes contribute to trait buffering, with more contribution from non-monotone curves. The direction of trait buffering is shown in the column on the right.

buffering, we decomposed ξ into the contribution from non-monotone and monotone GDRCs (Appendix E). These measures revealed that both monotone and non-monotone curves contribute to trait buffering (Figure 5D), with non-monotone GDRCs contributing a median of 80% to trait buffering.

Overall, traits with large non-monotone average burden effects have GDRCs that are buffered against one direction of the trait. Both of our proposed models (Figures 1E and 1F) contribute to produce the non-monotone average GDRCs, with non-monotone GDRCs contributing more strongly than monotone GDRCs.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

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.

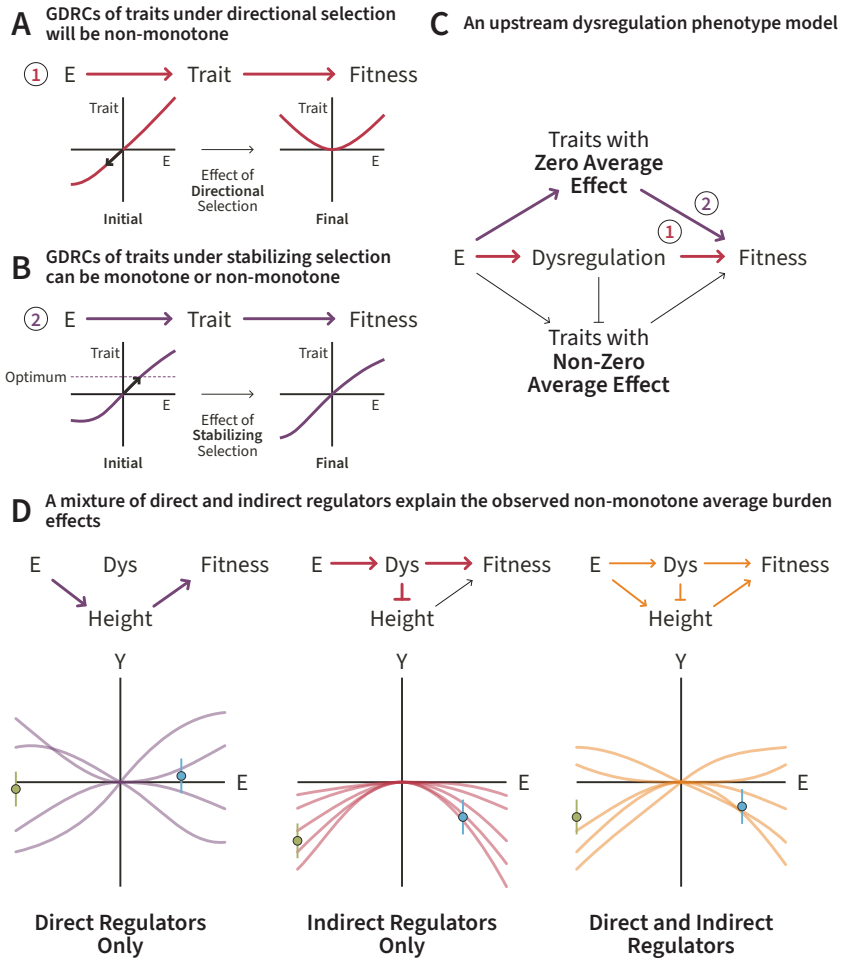

FIGURE 6. A theoretical model to explain trait buffering. A. GDRCs of traits under directional selection will become non-monotone over evolutionary time. **B.** GDRCs of traits under stabilizing selection will have no preference for GDRC shape. **C.** Traits with non-zero average burden effects have 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.

Two observations initially seem to imply that the complex traits with non-monotone average GDRCs are under directional selection. First, the GDRCs for these traits are buffered against a particular trait direction. Second, the non-monotone average GDRC suggests that any variant should have an average 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 that is under negative selection (Figure 6A). Then, selection will tune the expression of the gene until the trait is minimized, resulting in non-monotone GDRCs where both deletions and duplica-

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

tions increase the trait. Thus, the non-monotone average GDRC and trait buffering in the direction of the average burden effects suggests that the GDRCs of these complex traits are shaped by directional selection.

However, prior analyses using GWAS data for many of these traits suggest that they are under a different mode of selection – stabilizing selection [43–47]. Additionally, we do not see a non-zero average effect in GWAS data (Figure 1C, Appendix F), inconsistent with directional selection.

Stabilizing selection affects GDRCs differently than 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, and we do not 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 average GDRCs are downstream of an unobserved trait under negative 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 if 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 via 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 regulation or indirect regulation via dysregulation, we can still have predominantly monotone GDRCs but have a consistent bias toward one direction of the focal trait.

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 that acts via dysregulation can only affect the focal trait in one direction because it has a non-monotone GDRC, while 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 believe that this unobserved latent phenotype is a measure of dysregulation because the downstream traits with large non-monotone average burden effects are readouts of general dysfunction. The

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

traits with significant non-monotone average effects (Table F.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 or 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 [48, 49]. Age is associated with general dysregulation and disease [50]. Strong predictors of biological age in predictive models include FEV1, cystatin C, HbA1c, alkaline phosphatase, and hand grip strength [48, 49], 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. 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. The shape of GDRCs may thus be molded by the effects of directional selection.

Implications of the observed characteristics of gene dosage response curves

In a “linear world” where dosage has a linear relationship with trait, ascertaining any point on the GDRC along the dosage spectrum would carry all the information about the gene’s effect on 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 [51]. The shape of the GDRC implied by the burden estimates can also be useful prior for fine-mapping causal variants [52] and linking variants to genes [53–57].

The assumption of linearity in TWAS [5, 6, 58] and Mendelian randomization [59] approaches may cause estimated effect sizes to be close to 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.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

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

Moving from genetic associations towards 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 what 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 trait. 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 [60, 61]. 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 [62–65] and experimental approaches [66–68] 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 [30], are challenging to integrate into the GDRC framework. Accurate computational effect prediction [69] or experimental approaches such as deep mutational scanning [70] 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 [71]. 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 [67] will be necessary to estimate full 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.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

Methods

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 burden tests allowed us to compare effect sizes across the variant classes. We performed burden test using synonymous variants as a negative control to estimate any residual confounding, which we detail in Appendix F. We improved upon previous LoF burden approaches by using variable frequency filters based on gene constraint (Appendix F). 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 have the advantage of more accurate copy number estimation, higher resolution of CNV breakpoints, and better detection of rare CNVs [72].

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 use the regression with summary statistics (RSS) likelihood [73], 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 catalogued in Appendix D.

Individuals

We followed a protocol similar to a prior analysis of rare LoF variants in the UKB [30]. We use all 460K individuals with WES data in the UKB data. This is slightly larger than the fraction used by Backman et al., who focused on 430K individuals with genetic similarity to the EUR “superpopulation” as defined by the 1000 Genomes Project. Although relatedness and population structure can introduce environmental confounding, our pilot analyses with synonymous variants (Appendix F) indicated minimal levels of uncorrected confounding.

Phenotypes

We curated 410 continuous phenotypes in the UKB on which to perform burden tests, of which 91 were used for the main analysis in the paper. We required phenotypes to have measurements in at least 40K individuals. We also excluded any traits where more than 50% of the instances match the mode of the trait. This is to remove quasi-categorical traits, and is more relaxed compared to Backman et al., who exclude traits with a cutoff of 20%. For phenotypes that were collected more than once in returning visits, we use the first instance of the observation for each individual. Backman et al. 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.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

Separately, we acquired access to the full 500K baseline metabolomics data from the nuclear magnetic resonance (NMR) metabolomics data set generated by Nightingale Health PLC [74].

Genotypes

We included 15 genotyping principal components (PCs), estimated previously in the UKB [75]. Backman et al. 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 (Appendix F) supported this choice.

Following the lead of Backman et al., 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 [76].

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 [77] with the LOFTEE plugin [78]. 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., we use all LoF sites with MAF less than 1% but require that the misannotation probability be less than 10%. The misannotation probability [61] 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 [61]. 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 (Appendix F).

CNVs were previously called in the UKB using Illumina's DRAGEN software [31, 79]. 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. We observed that some samples had a large amount of genome-wide copy gain or loss, which is likely a genotyping error (Appendix F). To account for this, we removed any sample with greater than 10 Mbp of deleted or duplicated genome sequence.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

We used GENCODE version 39 to define gene intervals [80]. We excluded genes on the Y chromosome and the mitochondrial genome.

Burden Tests

We used REGENIE to perform gene-level burden tests [32]. 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.

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.

Monotonicity Model

We modeled the standardized trait of interest $\mathbf{Y}$ in N individuals as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\gamma}_1 + \mathbf{Z}\boldsymbol{\gamma}_2 + \boldsymbol{\epsilon}\,,$$

where $\mathbf{X}$ and $\mathbf{Z}$ are the burden genotypes from LoF variants and duplications respectively. The error $\boldsymbol{\epsilon}$ 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:

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

This allowed us to model the quantities of interest. The average effect of deletions and duplications is represented by $\overline{\gamma}_1\mathbf{1}$ and $\overline{\gamma}_2\mathbf{1}$ 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,

$$\phi \overset{\triangle}{\propto} -\mathbb{E}\left[\gamma_{1j}\gamma_{2j}\right]\,.$$

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.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

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 utilized the RSS likelihood [73]. An extensive description of the model is provided in Appendix B.

We developed unbiased estimators of $\overline{\gamma}_1$ and $\overline{\gamma}_2$. Since the monotonicity 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 Appendices C and D respectively.

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 [42], which is a flexible multivariate prior. We continued to use the RSS likelihood to model the observed effect sizes. MASH uses a mixture of K multivariate normal distributions to model the latent space:

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

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. Since the RSS likelihood necessarily induces dependencies between the observations via linkage disequilibrium (LD), we implemented a stochastic approximation expectation maximization (SAEM) algorithm to optimize $\boldsymbol{\pi}$ [81, 82]. An alternative approach to this model has been proposed that uses variational inference [83–85].

As we defined it, the amount of directional 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 Appendix E. We used the posterior distribution over $\boldsymbol{\gamma}$ to estimate

$$\xi \triangleq \mathbb{E}\left[\log_2\left(\frac{3}{2}\right)\gamma_{1j} + \gamma_{2j} \,\middle|\, \hat{\boldsymbol{\gamma}}\right] ,$$

which corresponds to the posterior expectation of the dot product given the observed data. To determine how much is contributed to ξ by monotone or non-monotone genes, we conditioned on the type of the

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

GDRC:

$$\xi_{\mathrm{nm}} = \mathbb{E}\left[\log_2\left(\frac{3}{2}\right)\gamma_{1j} + \gamma_{2j} \,\middle|\, \hat{\boldsymbol{\gamma}}, \gamma_{1j}\gamma_{2j} > 0\right]$$
$$\xi_{\mathrm{m}} = \mathbb{E}\left[\log_2\left(\frac{3}{2}\right)\gamma_{1j} + \gamma_{2j} \,\middle|\, \hat{\boldsymbol{\gamma}}, \gamma_{1j}\gamma_{2j} < 0\right] .$$

Using the law of total expectation, we can decompose ξ into the contribution from non-monotone and monotone GDRCs as

$$\xi = \xi_{\mathrm{nm}}\mathbb{P}\left(\gamma_{1j}\gamma_{2j} > 0\right) + \xi_{\mathrm{m}}\mathbb{P}\left(\gamma_{1j}\gamma_{2j} < 0\right) .$$

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 can be found in Appendix E.

Data Analysis

Throughout, we use unbiased maximum likelihood estimates (MLEs) for the duplication burden estimates when they are displayed in the paper, which account for LD 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 (Figures 1C and F.8). The collection and analysis of these data is described in Appendix F. We fit a locally estimated scatterplot smoothing (LOESS) curve to visualize the conditional mean of the derived allele effect given the derived allele frequency.

In Figure 2, top gene-trait pairs were ascertained with $|z| > 3$ for both LoF and duplication burden estimates. Gene-trait pairs were then ordered by $\hat{\gamma}^2_{\mathrm{LoF}}$ since we are more confident in the LoF estimate, and the top trait was chosen per gene. Thus, each gene was only represented once in the plot. Then, we ordered gene-trait pairs by $|z_{\mathrm{LoF}} z_{\mathrm{Dup}}|$ and chose the top hit per trait. By doing so, each trait also appears only once in the plot.

We inferred a distribution of effects for ϕ across traits for Figure 3B using a hierarchical model that was fit using empirical Bayes. The model is described in Appendix F.

In Figure 4A, top hits were ascertained as gene-trait pairs with $|z_{\mathrm{LoF}}| > 2$ and $|z_{\mathrm{Dup}}| > 2$ for traits with significant average burden effects (Table F.1). Then, to plot polarized Z scores, we multiplied the Z score by the sign of the average burden effect. This means that genes with positive burden effect estimates are in the same direction as the average burden effect. Since both the LoF and duplication burden effects are noisy estimates, we use the first PC of the Z scores to represent the trend of the data. The data was not centered or scaled before running PCA.

We estimated confident separation from the diagonal line in Figure 5B using the local false sign rate [86]. This rate is calculated using the number of posterior samples that fall on either side of the diagonal

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

for a given gene. A rate of less than 0.05 implies that less than 5% of the posterior density is on one side of the diagonal. Figure 5C represents the sample plot as Figure 1B. Points are colored based on the sign of ξ and the local false sign rate. Traits with significant average burden effects (Table F.1) were used in Figure 5D. The proportion of contribution from non-monotone GDRCs was estimated as

$$\frac{\xi_{\mathrm{nm}}\mathbb{P}\left(\gamma_{1j}\gamma_{2j} > 0\right)}{\xi_{\mathrm{m}}\mathbb{P}\left(\gamma_{1j}\gamma_{2j} < 0\right)}.$$

Data Availability

All genetic and health data was acquired from the UK Biobank, a biomedical database containing information from half a million UK participants (https://www.ukbiobank.ac.uk/). Data for genetic association analysis of continuous traits was acquired under application 52374. Data for genetic association analysis of metabolite traits was acquired under application 30418. These data are available upon application to the UK Biobank. All summary statistics generated from genetic association analysis and other processed data files are deposited in Zenodo at https://doi.org/10.5281/zenodo.13852455.

Code Availability

Scripts used to process and analyze data are deposited in Zenodo at https://doi.org/10.5281/zenodo.13864152. Scripts were executed either on the UK Biobank 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 would like to 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. This research has been conducted using data from the UK Biobank (https://www.ukbiobank.ac.uk/), a major biomedical database. We would like to 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 would like to 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 Stanford Biology Department's graduate student assistantship. This work was supported by the National Institutes of Health (R01HG011432 and R01HG008140 to J.K.P.).

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

Author Contributions

N.M. Methodology, Software, Formal Analysis, Data Curation, Writing – Original Draft, Visualization. **C.J.S.** Methodology, Software, Writing – Original Draft, Visualization. **H.Z.** Methodology, Software, Writing – Original Draft, Visualization. **J.P.S.** Conceptualization, Methodology, Writing – Original Draft, Supervision. **J.K.P.** Conceptualization, Methodology, Writing – Original Draft, Supervision, Project Administration, Funding Acquisition.

References

[1] Melina Claussnitzer et al. "A brief history of human disease genetics". *Nature* 577.7789 (Jan. 2020), pp. 179–189. DOI: 10.1038/s41586-019-1879-7.

[2] Elliot Sollis et al. "The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource". *Nucleic Acids Research* 51.D1 (Jan. 2023), pp. D977–D985. DOI: 10.1093/nar/gkac1010.

[3] Matthew T. Maurano et al. "Systematic Localization of Common Disease-Associated Variation in Regulatory DNA". *Science* 337.6099 (Sept. 2012), pp. 1190–1195. DOI: 10.1126/science.1222794.

[4] Anshul Kundaje et al. "Integrative analysis of 111 reference human epigenomes". *Nature* 518.7539 (Feb. 2015), pp. 317–330. DOI: 10.1038/nature14248.

[5] Eric R. Gamazon et al. "A gene-based association method for mapping traits using reference transcriptome data". *Nature Genetics* 47.9 (Aug. 2015), pp. 1091–1098. DOI: 10.1038/ng.3367.

[6] Alexander Gusev et al. "Integrative approaches for large-scale transcriptome-wide association studies". *Nature Genetics* 48.3 (Feb. 2016), pp. 245–252. DOI: 10.1038/ng.3506.

[7] Farhad Hormozdiari et al. "Colocalization of GWAS and eQTL Signals Detects Target Genes". *The American Journal of Human Genetics* 99.6 (Dec. 2016), pp. 1245–1260. DOI: 10.1016/j.ajhg.2016.10.003.

[8] Wouter Meuleman et al. "Index and biological spectrum of human DNase I hypersensitive sites". *Nature* 584.7820 (July 2020), pp. 244–251. DOI: 10.1038/s41586-020-2559-3.

[9] Silvia Stingele et al. "Global analysis of genome, transcriptome and proteome reveals the response to aneuploidy in human cells". *Molecular Systems Biology* 8.1 (Sept. 2012), p. 608. DOI: 10.1038/msb.2012.40.

[10] Terry Hassold and Patricia Hunt. "To err (meiotically) is human: the genesis of human aneuploidy". *Nature Reviews Genetics* 2.4 (Apr. 2001), pp. 280–291. DOI: 10.1038/35066065.

[11] C D Lehman et al. "Trisomy 13 syndrome: prenatal US findings in a review of 33 cases." *Radiology* 194.1 (Jan. 1995), pp. 217–222. DOI: 10.1148/radiology.194.1.7997556.

[12] Anna Cereda and John C. Carey. "The trisomy 18 syndrome". *Orphanet Journal of Rare Diseases* 7.1 (Oct. 2012), p. 81. DOI: 10.1186/1750-1172-7-81.

[13] Marilyn J. Bull. "Down Syndrome". *New England Journal of Medicine* 382.24 (June 2020), pp. 2344–2352. DOI: 10.1056/NEJMra1706537.

[14] Xin Shao et al. "Copy number variation is highly correlated with differential gene expression: a pan-cancer study". *BMC medical genetics* 20.1 (Nov. 2019), p. 175. DOI: 10.1186/s12881-019-0909-5.

[15] Peter H. Sudmant et al. "Global diversity, population stratification, and selection of human copy-number variation". *Science* 349.6253 (Sept. 2015), aab3761. DOI: 10.1126/science.aab3761.

[16] Haley J. Abel et al. "Mapping and characterization of structural variation in 17,795 human genomes". *Nature* 583.7814 (May 2020), pp. 83–89. DOI: 10.1038/s41586-020-2371-0.

[17] Ryan L. Collins et al. "A structural variation reference for medical and population genetics". *Nature* 581.7809 (May 2020), pp. 444–451. DOI: 10.1038/s41586-020-2287-8.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

[18] Mohamed A. Almarri et al. "Population Structure, Stratification, and Introgression of Human Structural Variation". *Cell* 182.1 (July 2020), 189–199.e15. DOI: 10.1016/j.cell.2020.05.024.

[19] Andrew Dauber et al. "Genome-wide Association of Copy-Number Variation Reveals an Association between Short Stature and the Presence of Low-Frequency Genomic Deletions". *The American Journal of Human Genetics* 89.6 (Dec. 2011), pp. 751–759. DOI: 10.1016/j.ajhg.2011.10.014.

[20] Eleanor Wheeler et al. "Genome-wide SNP and CNV analysis identifies common and low-frequency variants associated with severe early-onset obesity". *Nature Genetics* 45.5 (Apr. 2013), pp. 513–517. DOI: 10.1038/ng.2607.

[21] Aurélien Macé et al. "CNV-association meta-analysis in 191,161 European adults reveals new loci associated with anthropometric traits". *Nature Communications* 8.1 (Sept. 2017), p. 744. DOI: 10.1038/s41467-017-00556-x.

[22] Christian R. Marshall et al. "Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects". *Nature Genetics* 49.1 (Nov. 2016), pp. 27–35. DOI: 10.1038/ng.3725.

[23] Matthew Aguirre, Manuel A. Rivas, and James Priest. "Phenome-wide Burden of Copy-Number Variation in the UK Biobank". *The American Journal of Human Genetics* 105.2 (Aug. 2019), pp. 373–383. DOI: 10.1016/j.ajhg.2019.07.001.

[24] Chiara Auwerx et al. "The individual and global impact of copy-number variants on complex human traits". *The American Journal of Human Genetics* 109.4 (Apr. 2022), pp. 647–668. DOI: 10.1016/j.ajhg.2022.02.010.

[25] Margaux L. A. Hujoel et al. "Influences of rare copy-number variation on human complex traits". *Cell* 185.22 (Oct. 2022), 4233–4248.e27. DOI: 10.1016/j.cell.2022.09.028.

[26] Ryan L. Collins et al. "A cross-disorder dosage sensitivity map of the human genome". *Cell* 185.16 (Aug. 2022), 3041–3055.e25. DOI: 10.1016/j.cell.2022.06.036.

[27] Chiara Auwerx et al. *Copy-number variants as modulators of common disease susceptibility*. Aug. 2023. DOI: 10.1101/2023.07.31.23293408.

[28] Chia-Yen Chen et al. "The impact of rare protein coding genetic variation on adult cognitive function". *Nature Genetics* 55.6 (June 2023), pp. 927–938. DOI: 10.1038/s41588-023-01398-8.

[29] Yunfeng Huang et al. "Rare genetic variants impact muscle strength". *Nature Communications* 14.1 (June 2023), p. 3449. DOI: 10.1038/s41467-023-39247-1.

[30] Joshua D. Backman et al. "Exome sequencing and analysis of 454,787 UK Biobank participants". *Nature* 599.7886 (Oct. 2021), pp. 628–634. DOI: 10.1038/s41586-021-04103-z.

[31] Shuwei Li et al. *Whole-genome sequencing of half-a-million UK Biobank participants*. Dec. 2023. DOI: 10.1101/2023.12.06.23299426.

[32] Joelle Mbatchou et al. "Computationally efficient whole-genome regression for quantitative and binary traits". *Nature Genetics* 53.7 (May 2021), pp. 1097–1103. DOI: 10.1038/s41588-021-00870-7.

[33] Sahin Naqvi et al. "Precise modulation of transcription factor levels identifies features underlying dosage sensitivity". *Nature Genetics* 55.5 (May 2023), pp. 841–851. DOI: 10.1038/s41588-023-01366-2.

[34] Júlia Domingo et al. *Non-linear transcriptional responses to gradual modulation of transcription factor dosage*. Mar. 2024. DOI: 10.1101/2024.03.01.582837.

[35] Sébastien Jacquemont et al. "Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus". *Nature* 478.7367 (Aug. 2011), pp. 97–102. DOI: 10.1038/nature10406.

[36] Christelle Golzio et al. "KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant". *Nature* 485.7398 (May 2012), pp. 363–367. DOI: 10.1038/nature11091.

[37] Tülay Güran. "Latest Insights on the Etiology and Management of Primary Adrenal Insufficiency in Children". *J Clin Res Pediatr Endocrinol* 9.60 (2017), pp. 9–22. DOI: 10.4274/jcrpe.2017.S002.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

[38] Caleb B. Cantrell and Shamim S. Mohiuddin. *Biochemistry, Ketone Metabolism*. StatPearls Publishing, Treasure Island (FL), Mar. 2020.

[39] Agata Juszczak, Damian Morris, and Ashley Grossman. "Cushing's Syndrome". *Endotext*. Ed. by Kenneth R. Feingold et al. South Dartmouth (MA): MDText.com, Inc., 2000.

[40] Juliette Emmerich et al. "Cushing's Syndrome: Development of Highly Potent and Selective CYP11B1 Inhibitors of the (Pyridylmethyl)pyridine Type". *Journal of Medicinal Chemistry* 56.15 (Aug. 2013), pp. 6022–6032. DOI: 10.1021/jm400240r.

[41] Vladimir Stojanovic and Sherri Ihle. "Role of beta-hydroxybutyric acid in diabetic ketoacidosis: A review". *The Canadian Veterinary Journal* 52.4 (Apr. 2011), pp. 426–430.

[42] Sarah M. Urbut et al. "Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions". *Nature Genetics* 51.1 (Nov. 2018), pp. 187–195. DOI: 10.1038/s41588-018-0268-8.

[43] Yuval B. Simons et al. "A population genetic interpretation of GWAS findings for human quantitative traits". *PLOS Biology* 16.3 (Mar. 2018), e2002985. DOI: 10.1371/journal.pbio.2002985.

[44] Guy Sella and Nicholas H. Barton. "Thinking About the Evolution of Complex Traits in the Era of Genome-Wide Association Studies". *Annual Review of Genomics and Human Genetics* 20.1 (Aug. 2019), pp. 461–493. DOI: 10.1146/annurev-genom-083115-022316.

[45] Yuval B. Simons et al. *Simple scaling laws control the genetic architectures of human complex traits*. Oct. 2022. DOI: 10.1101/2022.10.04.509926.

[46] E. Koch et al. *Genetic association data are broadly consistent with stabilizing selection shaping human common diseases and traits*. July 2024. DOI: 10.1101/2024.06.19.599789.

[47] Roshni A. Patel et al. *Conditional frequency spectra as a tool for studying selection on complex traits in biobanks*. June 2024. DOI: 10.1101/2024.06.15.599126.

[48] Mei Sum Chan et al. "A Biomarker-based Biological Age in UK Biobank: Composition and Prediction of Mortality and Hospital Admissions". *The Journals of Gerontology: Series A* 76.7 (July 2021), pp. 1295–1302. DOI: 10.1093/gerona/glab069.

[49] Sergiy Libert, Alex Chekholko, and Cynthia Kenyon. *A mathematical model that predicts human biological age from physiological traits identifies environmental and genetic factors that influence aging*. Oct. 2023. DOI: 10.1101/2022.04.14.488358.

[50] Carlos López-Otín et al. "The Hallmarks of Aging". *Cell* 153.6 (June 2013), pp. 1194–1217. DOI: 10.1016/j.cell.2013.05.039.

[51] Ralf Schmidt et al. "CRISPR activation and interference screens decode stimulation responses in primary human T cells". *Science* 375.6580 (Feb. 2022), eabj4008. DOI: 10.1126/science.abj4008.

[52] Anna Hutchinson, Jennifer Asimit, and Chris Wallace. "Fine-mapping genetic associations". *Human Molecular Genetics* 29.R1 (Sept. 2020), R81–R88. DOI: 10.1093/hmg/ddaa148.

[53] Tuuli Lappalainen and Daniel G. MacArthur. "From variant to function in human disease genetics". *Science* 373.6562 (Sept. 2021), pp. 1464–1468. DOI: 10.1126/science.abi8207.

[54] Joseph Nasser et al. "Genome-wide enhancer maps link risk variants to disease genes". *Nature* 593.7858 (Apr. 2021), pp. 238–243. DOI: 10.1038/s41586-021-03446-x.

[55] Noah J Connally et al. "The missing link between genetic association and regulatory function". *eLife* 11 (Dec. 2022). Ed. by Jonathan Flint and Molly Przeworski, e74970. DOI: 10.7554/eLife.74970.

[56] Steven Gazal et al. "Combining SNP-to-gene linking strategies to identify disease genes and assess disease omnigenicity". *Nature Genetics* 54.6 (June 2022), pp. 827–836. DOI: 10.1038/s41588-022-01087-y.

[57] Andreas R. Gschwind et al. *An encyclopedia of enhancer-gene regulatory interactions in the human genome*. Nov. 2023. DOI: 10.1101/2023.11.09.563812.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

[58] Nicholas Mancuso et al. "Probabilistic fine-mapping of transcriptome-wide association studies". *Nature Genetics* 51.4 (Mar. 2019), pp. 675–682. DOI: 10.1038/s41588-019-0367-1.

[59] Zhihong Zhu et al. "Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets". *Nature Genetics* 48.5 (Mar. 2016), pp. 481–487. DOI: 10.1038/ng.3538.

[60] Hakhamanesh Mostafavi et al. "Systematic differences in discovery of genetic effects on gene expression and complex traits". *Nature Genetics* (Oct. 2023), pp. 1–10. DOI: 10.1038/s41588-023-01529-1.

[61] Tony Zeng et al. "Bayesian estimation of gene constraint from an evolutionary model with gene features". *Nature Genetics* (July 2024), pp. 1–12. DOI: 10.1038/s41588-024-01820-9.

[62] David R. Kelley et al. "Sequential regulatory activity prediction across chromosomes with convolutional neural networks". *Genome Research* 28.5 (Mar. 2018), pp. 739–750. DOI: 10.1101/gr.227819.117.

[63] Vikram Agarwal and Jay Shendure. "Predicting mRNA Abundance Directly from Genomic Sequence Using Deep Convolutional Neural Networks". *Cell Reports* 31.7 (May 2020). DOI: 10.1016/j.celrep.2020.107663.

[64] Žiga Avsec et al. "Effective gene expression prediction from sequence by integrating long-range interactions". *Nature Methods* 18.10 (Oct. 2021), pp. 1196–1203. DOI: 10.1038/s41592-021-01252-x.

[65] Shiron Drusinsky, Sean Whalen, and Katherine S. Pollard. *Deep-learning prediction of gene expression from personal genomes*. July 2024. DOI: 10.1101/2024.07.27.605449.

[66] Atray Dixit et al. "Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens". *Cell* 167.7 (Dec. 2016), 1853–1866.e17. DOI: 10.1016/j.cell.2016.11.038.

[67] Ryan Tewhey et al. "Direct Identification of Hundreds of Expression-Modulating Variants using a Multiplexed Reporter Assay". *Cell* 165.6 (June 2016), pp. 1519–1529. DOI: 10.1016/j.cell.2016.04.027.

[68] Martin Kircher et al. "Saturation mutagenesis of twenty disease-associated regulatory elements at single base-pair resolution". *Nature Communications* 10.1 (Aug. 2019), p. 3583. DOI: 10.1038/s41467-019-11526-w.

[69] Jun Cheng et al. "Accurate proteome-wide missense variant effect prediction with AlphaMissense". *Science* 381.6664 (Sept. 2023), eadg7492. DOI: 10.1126/science.adg7492.

[70] Douglas M. Fowler and Stanley Fields. "Deep mutational scanning: a new style of protein science". *Nature Methods* 11.8 (July 2014), pp. 801–807. DOI: 10.1038/nmeth.3027.

[71] GTEx Consortium. "Genetic effects on gene expression across human tissues". *Nature* 550.7675 (Oct. 2017), pp. 204–213. DOI: 10.1038/nature24277.

[72] Min Zhao et al. "Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives". *BMC Bioinformatics* 14.11 (Sept. 2013), S1. DOI: 10.1186/1471-2105-14-S11-S1.

[73] Xiang Zhu and Matthew Stephens. "Bayesian large-scale multiple regression with summary statistics from genome-wide association studies". *The Annals of Applied Statistics* 11.3 (Sept. 2017), pp. 1561–1592. DOI: 10.1214/17-AOAS1046.

[74] Heli Julkunen et al. "Atlas of plasma NMR biomarkers for health and disease in 118,461 individuals from the UK Biobank". *Nature Communications* 14.1 (Feb. 2023), p. 604. DOI: 10.1038/s41467-023-36231-7.

[75] Clare Bycroft et al. "The UK Biobank resource with deep phenotyping and genomic data". *Nature* 562.7726 (Oct. 2018), pp. 203–209. DOI: 10.1038/s41586-018-0579-z.

[76] Gad Abraham, Yixuan Qiu, and Michael Inouye. "FlashPCA2: principal component analysis of Biobank-scale genotype datasets". *Bioinformatics* 33.17 (Sept. 2017), pp. 2776–2778. DOI: 10.1093/bioinformatics/btx299.

[77] William McLaren et al. "The Ensembl Variant Effect Predictor". *Genome Biology* 17.1 (June 2016), p. 122. DOI: 10.1186/s13059-016-0974-4.

[78] Konrad J. Karczewski et al. "The mutational constraint spectrum quantified from variation in 141,456 humans". *Nature* 581.7809 (May 2020), pp. 434–443. DOI: 10.1038/s41586-020-2308-7.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

[79] Sairam Behera et al. "Comprehensive genome analysis and variant detection at scale using DRAGEN". *Nature Biotechnology* (Oct. 2024), pp. 1–15. DOI: 10.1038/s41587-024-02382-1.

[80] Adam Frankish et al. "GENCODE: reference annotation for the human and mouse genomes in 2023". *Nucleic Acids Research* 51.D1 (Jan. 2023), pp. D942–D949. DOI: 10.1093/nar/gkac1071.

[81] Bernard Delyon, Marc Lavielle, and Eric Moulines. "Convergence of a Stochastic Approximation Version of the EM Algorithm". *The Annals of Statistics* 27.1 (Feb. 1999), pp. 94–128.

[82] Estelle Kuhn and Marc Lavielle. "Coupling a stochastic approximation version of EM with an MCMC procedure". *ESAIM: Probability and Statistics* 8 (Aug. 2004), pp. 115–131. DOI: 10.1051/ps:2004007.

[83] Jeffrey P. Spence et al. *A flexible modeling and inference framework for estimating variant effect sizes from GWAS summary statistics*. Apr. 2022. DOI: 10.1101/2022.04.18.488696.

[84] Fabio Morgante et al. "A flexible empirical Bayes approach to multivariate multiple regression, and its improved accuracy in predicting multi-tissue gene expression from genotypes". *PLOS Genetics* 19.7 (July 2023), e1010539. DOI: 10.1371/journal.pgen.1010539.

[85] Deborah Kunkel et al. *Improving polygenic prediction from summary data by learning patterns of effect sharing across multiple phenotypes*. May 2024. DOI: 10.1101/2024.05.06.592745.

[86] Matthew Stephens. "False discovery rates: a new deal". *Biostatistics* (Oct. 2016), kxw041. DOI: 10 . 1093 / biostatistics/kxw041.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

Appendices

A Gene Dosage Response Curves

Genome-wide burden

We retrieved genome-wide burden effect estimates from work by Auwerx et al., where burden was defined as the number of genes affected by deletions or duplications [1]. We subset to traits where both the genome-wide deletion and duplication burden effects had a p-value of less than 0.05. In all cases, the deletion and duplication burden effect is in the same direction.

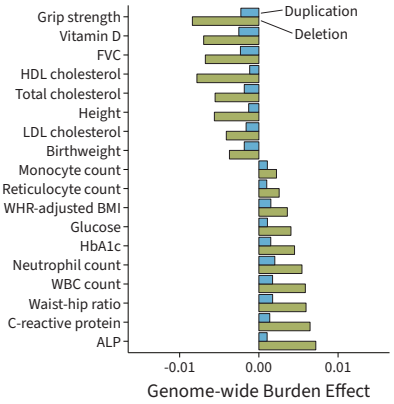

Figure A.1. Genome-wide burden data, reprocessed from work by Auwerx et al. [1]. These effects have a p-value of less than 0.05. The deletion or duplication effect is the effect of increasing numbers of deleted/duplicated genes on the trait.

Scale of the dosage and trait

The scale and shape of the curve depend on how gene expression and the trait are measured. We assume that expression is measured on a logarithmic scale relative to the population mean. This allows us to define the expression effect of loss-of-function (LoF) and duplication variants on a logarithmic scale such that expression-decreasing variants have negative expression values and expression-increasing variants have positive expression values. A heterozygous LoF genotype thus has an expression effect of $\log_2\left(\frac{1}{2}\right) = -1$, and a duplication resulting in a copy number of three has an expression effect of $\log_2\left(\frac{3}{2}\right) \approx 0.59$. For traits, we chose to use an inverse rank normal transform. By centering the trait, we implicitly define trait-decreasing variants as having negative effects and trait-increasing variants as having positive effects.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

The consequence of centering both expression and trait is that the gene dosage response curve (GDRC) always passes through the origin.

Composition of monotone curves

Let $\gamma : E \mapsto Y$ be the GDRC for a gene. Suppose that the dosage effect percolates to the trait via a path, as diagrammed in Figure 4C. Consider a path consisting of two edges with associated functions $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$. The GDRC along this path will then be $Y = \gamma(E) = f(g(E)) = (f \circ g)(E)$. Here, f and g can are the curves of individual arrows along the path.

Now, suppose that g is a monotone increasing function. That is, $x \leq y$ implies that $g(x) \leq g(y)$ for any $x, y \in \mathbb{R}$. Similarly, suppose that f is a monotone increasing function. Then, $x \leq y$ implies that $g(x) \leq g(y)$, which implies that $f(g(x)) \leq f(g(y))$. Thus, the composition $(f \circ g)$ is also monotone increasing. Similar arguments for monotone decreasing functions or pairs of increasing and decreasing functions demonstrate that the composition of monotone functions is also monotone.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

B Monotonicity Model

Here, we endeavor to jointly model burden summary statistics from different points on the dosage spectrum. Burden summary statistics are often reported as an effect size and a standard error from regression. We use hierarchical models to account for the sampling error and pool information across genes.

Let $\mathbf{Y} \in \mathbb{R}^N$ be a standardized phenotype of interest, measured in N individuals. Let $\mathbf{X}, \mathbf{Z} \in \mathbb{R}^{N \times M}$ represent the burden genotype matrices for two variant classes respectively. By variant class, we mean a set of variants with a dosage effect in the same direction, such as deletions, duplications, or LoF variants. We assume that there are M genes that are polymorphic for both classes of variants. The genotypes are encoded as copies of the dosage-perturbing allele.

We use the following linear system to model the phenotype:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\gamma}_1 + \mathbf{Z}\boldsymbol{\gamma}_2 + \boldsymbol{\epsilon}\,.$$

Here, $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \in \mathbb{R}^M$ are the unobserved, per-allele effect sizes of perturbations of each gene on the phenotype. The residual error is represented by $\boldsymbol{\epsilon} \in \mathbb{R}^N$ and is assumed to be drawn from an isotropic normal distribution,

$$\boldsymbol{\epsilon} \sim \mathcal{N}\left(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}\right)\,.$$

Distribution of Effect Sizes

We model the latent effect sizes as coming from an uncentered normal distribution. Effect sizes between genes are assumed to be independent, but effect sizes for the same gene covary between the variant classes. That is to say, the effect of a deletion or LoF variant is assumed to carry some information about the effect of a duplication of the same gene. This covariance structure is represented with diagonal matrices $\boldsymbol{\Sigma}_{11}$, $\boldsymbol{\Sigma}_{12}$, and $\boldsymbol{\Sigma}_{22}$.

$$\begin{bmatrix} \boldsymbol{\gamma}_1 \\ \boldsymbol{\gamma}_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \overline{\boldsymbol{\gamma}}_1 \\ \overline{\boldsymbol{\gamma}}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12}^\top & \boldsymbol{\Sigma}_{22} \end{bmatrix} \right)$$

Suppose we collect the block matrices above into $\boldsymbol{\gamma}, \overline{\boldsymbol{\gamma}} \in \mathbb{R}^{2M}$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{2M \times 2M}$. Then, we can succinctly state that

$$\boldsymbol{\gamma} \sim \mathcal{N}\left(\overline{\boldsymbol{\gamma}}, \boldsymbol{\Sigma}\right)\,.$$

Let $\boldsymbol{\gamma}_{\cdot j} \in \mathbb{R}^2$ represent the latent effect sizes for the jth gene. For instance, the first coordinate may represent the LoF effect, and the second coordinate may represent the duplication effect. Then, the marginal distribution is written as

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

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

Parameter of Interest

We consider a gene to have a "monotone" effect on trait if the dosage-reducing alleles have an opposite direction of effect compared to dosage-increasing alleles. To this end, we define the monotonicity ϕ as a value proportional to the negative uncentered second moment of the effect sizes,

$$\phi \overset{\triangle}{\propto} -\mathbb{E}\left[\gamma_{1j}\gamma_{2j}\right] .$$

Thus, a positive ϕ represents monotone behavior. To compare monotonicity across traits, we restrict its codomain to $[-1, 1]$. We do this using the Cauchy-Schwarz inequality, which guarantees that

$$\left|\mathbb{E}\left[\gamma_{1j}\gamma_{2j}\right]\right|^2 \leq \mathbb{E}\left[\gamma_{1j}^2\right]\mathbb{E}\left[\gamma_{2j}^2\right] ,$$

so that

$$\phi \overset{\triangle}{=} -\frac{\mathbb{E}\left[\gamma_{1j}\gamma_{2j}\right]}{\sqrt{\mathbb{E}\left[\gamma_{1j}^2\right]\mathbb{E}\left[\gamma_{2j}^2\right]}} = -\frac{\overline{\gamma}_1\overline{\gamma}_2 + \sigma_{12}}{\sqrt{\left(\overline{\gamma}_1^2 + \sigma_{11}^2\right)\left(\overline{\gamma}_2^2 + \sigma_{22}^2\right)}} .$$

When the effect sizes are centered (that is, $\overline{\gamma}_1 = \overline{\gamma}_2 = 0$), this is equivalent to the negative correlation coefficient,

$$\phi = -\rho = -\frac{\sigma_{12}}{\sigma_{11}\sigma_{22}} .$$

Distribution of Genotypes

We begin by assuming that there is no linkage disequilibrium (LD) between the different variant classes. This is reasonable to assume because different variants arise through different mutational processes, and burden genotypes represent aggregates of multiple rare variants which generally have negligible LD [2]. However, this assumption does not apply across genes within a variant class. Within duplications, for instance, large duplications consisting of multiple genes will induce LD between the burden genotypes.

We assume that the M burden genotypes are in Hardy-Weinberg equilibrium (HWE), which implies a binomial likelihood. Let p_j represent the allele frequency of the jth gene's dosage-perturbing allele for the first variant class and let q_j represent the allele frequency of the jth gene's dosage-perturbing allele for the second variant class. Without loss of generality, we assume that the genotypes are centered but not scaled:

$$\mathbf{X}_{ij} + 2p_j \overset{\text{iid}}{\sim} \text{Binomial}\left(2, p_j\right)$$
$$\mathbf{Z}_{ij} + 2q_j \overset{\text{iid}}{\sim} \text{Binomial}\left(2, q_j\right) .$$

There is no LD between variant classes,

$$\mathbb{E}\left[\mathbf{X}_{ij}\mathbf{Z}_{k\ell}\right] = 0 \quad \forall i, j, k, \ell .$$

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

However, there is LD between the same variant class. The correlation matrices $\mathbf{R}_1$ and $\mathbf{R}_2$ are used to represent this LD. Let $P_j = 2p_j(1-p_j)$ and $Q_j = 2q_j(1-q_j)$ represent the heterozygosity of the burden genotypes for the jth gene. Then,

$$\mathbb{E}\left[\mathbf{X}^\top\mathbf{X}\right] = \mathbf{P}^{\frac{1}{2}}\mathbf{R}_1\mathbf{P}^{\frac{1}{2}}$$
$$\mathbb{E}\left[\mathbf{Z}^\top\mathbf{Z}\right] = \mathbf{Q}^{\frac{1}{2}}\mathbf{R}_2\mathbf{Q}^{\frac{1}{2}}\,,$$

where

$$\mathbf{P} = \mathrm{diag}\left(\{P_j\}_{j=1}^M\right)$$
$$\mathbf{Q} = \mathrm{diag}\left(\{Q_j\}_{j=1}^M\right)\,.$$

Since LoF variants affect individual genes, and we restrict ourselves to analyzing rare variants, we assume that the LoF burden genotypes are independent and that there is no LD between them. That is, $\mathbf{R}_1 = \mathbf{I}$. Since deletions and duplications can span multiple genes, the burden genotypes are not independent.

Burden Regression

Burden test results are reported as summary statistics from regression between the burden genotype and the phenotype. Regression is performed on centered and scaled phenotypes. The marginal summary statistics are calculated based on the following model for each gene:

$$\mathbf{Y} = \mathbf{X}_{\cdot j}\gamma_{1j} + \boldsymbol{\epsilon}_1$$
$$\mathbf{Y} = \mathbf{Z}_{\cdot j}\gamma_{2j} + \boldsymbol{\epsilon}_2\,.$$

The effect size is approximately the following. Note that $\left(\mathbf{X}_{\cdot j}^\top\mathbf{X}_{\cdot j}\right)^{-1} \approx \frac{1}{NP_j}$ and $\left(\mathbf{Z}_{\cdot j}^\top\mathbf{Z}_{\cdot j}\right)^{-1} \approx \frac{1}{NQ_j}$, so

$$\hat{\gamma}_{1j} = \left(\mathbf{X}_{\cdot j}^\top\mathbf{X}_{\cdot j}\right)^{-1}\mathbf{X}_{\cdot j}^\top\mathbf{Y} \approx \frac{1}{NP_j}\mathbf{X}_{\cdot j}^\top\mathbf{Y}$$
$$\hat{\gamma}_{2j} = \left(\mathbf{Z}_{\cdot j}^\top\mathbf{Z}_{\cdot j}\right)^{-1}\mathbf{Z}_{\cdot j}^\top\mathbf{Y} \approx \frac{1}{NQ_j}\mathbf{Z}_{\cdot j}^\top\mathbf{Y}\,.$$

The standard error is approximately the following. Here, we assume that any individual gene explains a small fraction of the total variance. That is, $\sigma_{\epsilon_1}^2 \approx 1$ and $\sigma_{\epsilon_2}^2 \approx 1$. Thus,

$$\hat{s}_{1j} \triangleq \widehat{\mathrm{SE}}\left(\hat{\gamma}_{1j}\right) = \sqrt{\sigma_{\epsilon_1}^2\left(\mathbf{X}_{\cdot j}^\top\mathbf{X}_{\cdot j}\right)^{-1}} \approx \frac{1}{\sqrt{NP_j}}$$
$$\hat{s}_{2j} \triangleq \widehat{\mathrm{SE}}\left(\hat{\gamma}_{2j}\right) = \sqrt{\sigma_{\epsilon_2}^2\left(\mathbf{Z}_{\cdot j}^\top\mathbf{Z}_{\cdot j}\right)^{-1}} \approx \frac{1}{\sqrt{NQ_j}}\,.$$

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

Therefore, the Z scores are

$$z_{1j} \approx \frac{1}{\sqrt{NP_j}} \mathbf{X}_{\cdot j}^\top \mathbf{Y}$$
$$z_{2j} \approx \frac{1}{\sqrt{NQ_j}} \mathbf{Z}_{\cdot j}^\top \mathbf{Y} \, .$$

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

C Monotonicity Inference

Maximum Likelihood Estimation

For inference, we will use the summary statistics directly rather than the underlying genotype data. That is, we are given the estimated effect sizes ($\hat{\boldsymbol{\gamma}}_1$ and $\hat{\boldsymbol{\gamma}}_2$) and their standard errors ($\hat{\mathbf{s}}_1$ and $\hat{\mathbf{s}}_2$). To simplify notation, we represent the prior as

$$\boldsymbol{\gamma} \sim \mathcal{N}\left(\overline{\boldsymbol{\gamma}}\left(\boldsymbol{\theta}\right), \boldsymbol{\Sigma}\left(\boldsymbol{\omega}\right)\right) ,$$

where $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$ represent values used to parameterize the priors.

Next, the likelihood of the estimated effect sizes is defined using the approximate regression with summary statistics (RSS) likelihood [3]. Under this likelihood, the estimated effect sizes are patterned by the LD in the cohort. Furthermore, the standard error of the sampling distribution is used as an estimate of the dispersion around the mean. Zhu et al. showed that this likelihood asymptotically approaches the sampling distribution of the estimated effect sizes [3]. Let $\hat{\mathbf{S}}_1 = \operatorname{diag}\left(\{s_{1j}\}_{j=1}^{M}\right)$ and $\hat{\mathbf{S}}_2 = \operatorname{diag}\left(\{s_{2j}\}_{j=1}^{M}\right)$, and let $\hat{\mathbf{R}}_1$ and $\hat{\mathbf{R}}_2$ represent the in-sample correlation matrices of the two variant classes respectively. Then, in notation reflecting the RSS likelihood, we define

$$\hat{\mathbf{S}} \triangleq \begin{bmatrix} \hat{\mathbf{S}}_1 & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{S}}_2 \end{bmatrix}$$
$$\hat{\mathbf{R}} \triangleq \begin{bmatrix} \hat{\mathbf{R}}_1 & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{R}}_2 \end{bmatrix} .$$

Then, the RSS likelihood for the observed effect sizes is

$$\hat{\boldsymbol{\gamma}} \mid \boldsymbol{\gamma} \sim \mathcal{N}\left(\hat{\mathbf{S}}\hat{\mathbf{R}}\hat{\mathbf{S}}^{-1}\boldsymbol{\gamma}, \hat{\mathbf{S}}\hat{\mathbf{R}}\hat{\mathbf{S}}\right) .$$

Some genes are in perfect LD with each other. That is, the sample correlation matrix $\hat{\mathbf{R}}$ is not strictly positive definite. To improve numerical stability and to account for perfect LD, we project the data onto the linear subspace of dimension $L < M$ spanned by the LD matrix (that is, a projection orthogonal to the null space). Consider the following eigendecomposition, with a matrix with orthogonal columns of eigenvectors $\mathbf{U} \in \mathbb{R}^{M \times L}$ and a diagonal matrix of eigenvalues $\boldsymbol{\Lambda} \in \mathbb{R}^{L \times L}$,

$$\hat{\mathbf{S}}\hat{\mathbf{R}}\hat{\mathbf{S}} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\top} .$$

In practice, these are derived by dropping small eigenvalues from the numerical eigendecomposition of $\hat{\mathbf{S}}\hat{\mathbf{R}}\hat{\mathbf{S}}$. Rather than modeling the observations $\hat{\boldsymbol{\gamma}}$, we model the projected data,

$$\mathbf{U}^{\top}\hat{\boldsymbol{\gamma}} \mid \boldsymbol{\gamma} \sim \mathcal{N}\left(\mathbf{U}^{\top}\hat{\mathbf{S}}\hat{\mathbf{R}}\hat{\mathbf{S}}^{-1}\boldsymbol{\gamma}, \boldsymbol{\Lambda}\right) .$$

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

Note that under this model, the marginal likelihood of the estimates is

$$\mathbf{U}^\top \hat{\boldsymbol{\gamma}} \sim \mathcal{N}\left(\mathbf{U}^\top \hat{\mathbf{S}}\hat{\mathbf{R}}\hat{\mathbf{S}}^{-1}\overline{\boldsymbol{\gamma}}, \mathbf{U}^\top \hat{\mathbf{S}}\hat{\mathbf{R}}\hat{\mathbf{S}}^{-1}\boldsymbol{\Sigma}\hat{\mathbf{S}}^{-1}\hat{\mathbf{R}}\hat{\mathbf{S}}\mathbf{U} + \boldsymbol{\Lambda}\right) .$$

To simplify notation, let $\mathbf{A} = \mathbf{U}^\top \hat{\mathbf{S}}\hat{\mathbf{R}}\hat{\mathbf{S}}^{-1}$ and let $\mathbf{K} = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top + \boldsymbol{\Lambda}$. Then,

$$\mathbf{U}^\top \hat{\boldsymbol{\gamma}} \sim \mathcal{N}\left(\mathbf{A}\overline{\boldsymbol{\gamma}}, \mathbf{K}\right) .$$

The pseudoinverse $\mathbf{A}^+$ is a useful quantity in downstream derivations. Since the rotation of the data is defined, we can construct a pseudoinverse using components of the eigendecomposition. The pseudoinverse is

$$\mathbf{A}^+ = \hat{\mathbf{S}}^2\mathbf{U}\boldsymbol{\Lambda}^{-1} .$$

This pseudoinverse is a right pseudoinverse, as can be seen by

$$\begin{aligned}
\mathbf{A}\mathbf{A}^+ &= \mathbf{U}^\top \hat{\mathbf{S}}\hat{\mathbf{R}}\hat{\mathbf{S}}^{-1}\hat{\mathbf{S}}^2\mathbf{U}\boldsymbol{\Lambda}^{-1} \\
&= \mathbf{U}^\top \hat{\mathbf{S}}\hat{\mathbf{R}}\hat{\mathbf{S}}\hat{\mathbf{S}}^{-2}\hat{\mathbf{S}}^2\mathbf{U}\boldsymbol{\Lambda}^{-1} \\
&= \mathbf{U}^\top \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top \mathbf{U}\boldsymbol{\Lambda}^{-1} \\
&= \mathbf{I} .
\end{aligned}$$

The other properties of the pseudoinverse are readily confirmed using matrix algebra.

Inference

We maximize the marginal likelihood of the observed effect sizes with respect to the parameters of the model. The likelihood is

$$\begin{aligned}
\mathcal{L}\left(\boldsymbol{\theta}, \boldsymbol{\omega} \mid \hat{\boldsymbol{\gamma}}, \hat{\mathrm{s}}\right) &= \mathcal{N}\left(\mathbf{U}^\top \hat{\boldsymbol{\gamma}}\,; \mathbf{A}\overline{\boldsymbol{\gamma}}, \mathbf{K}\right) \\
&= (2\pi)^{-M} \left(\det \mathbf{K}\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left(\mathbf{U}^\top \hat{\boldsymbol{\gamma}} - \mathbf{A}\overline{\boldsymbol{\gamma}}\right)^\top \mathbf{K}^{-1}\left(\mathbf{U}^\top \hat{\boldsymbol{\gamma}} - \mathbf{A}\overline{\boldsymbol{\gamma}}\right)\right) .
\end{aligned}$$

The log likelihood is

$$\begin{aligned}
\ell\left(\boldsymbol{\theta}, \boldsymbol{\omega} \mid \hat{\boldsymbol{\gamma}}, \hat{\mathrm{s}}\right) &= \log \mathcal{L}\left(\boldsymbol{\gamma}, \boldsymbol{\omega} \mid \hat{\boldsymbol{\gamma}}, \hat{\mathrm{s}}\right) \\
&= -M \log\left(2\pi\right) - \frac{1}{2}\log\det \mathbf{K} - \frac{1}{2}\left(\mathbf{U}^\top \hat{\boldsymbol{\gamma}} - \mathbf{A}\overline{\boldsymbol{\gamma}}\right)^\top \mathbf{K}^{-1}\left(\mathbf{U}^\top \hat{\boldsymbol{\gamma}} - \mathbf{A}\overline{\boldsymbol{\gamma}}\right) .
\end{aligned}$$

We are interested in obtaining the maximum likelihood estimates $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\omega}}$ such that

$$\begin{aligned}
\hat{\boldsymbol{\theta}} &= \arg\max_{\boldsymbol{\theta}} \mathcal{L}\left(\boldsymbol{\theta}, \boldsymbol{\omega} \mid \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{s}}\right) \\
\hat{\boldsymbol{\omega}} &= \arg\max_{\boldsymbol{\omega}} \mathcal{L}\left(\boldsymbol{\theta}, \boldsymbol{\omega} \mid \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{s}}\right) .
\end{aligned}$$

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

For maximum likelihood estimation of $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$, we use natural gradient ascent [4]. This is an optimization approach that uses both first- and second-order information about the log likelihood via the gradient and Fisher information matrix. The gradient of the log likelihood function with respect to $\overline{\boldsymbol{\gamma}}$ is

$$\begin{aligned}\nabla_{\overline{\boldsymbol{\gamma}}}\,\ell &= \nabla_{\overline{\boldsymbol{\gamma}}}\left[-\frac{1}{2}\left(\mathbf{U}^\top\hat{\boldsymbol{\gamma}}-\mathbf{A}\overline{\boldsymbol{\gamma}}\right)^\top\mathbf{K}^{-1}\left(\mathbf{U}^\top\hat{\boldsymbol{\gamma}}-\mathbf{A}\overline{\boldsymbol{\gamma}}\right)\right]\\ &= \mathbf{A}^\top\mathbf{K}^{-1}\left(\mathbf{U}^\top\hat{\boldsymbol{\gamma}}-\mathbf{A}\overline{\boldsymbol{\gamma}}\right).\end{aligned}$$

Using the matrix chain rule, the derivative of the log likelihood with respect to one of the mean parameters is

$$\frac{\partial\ell}{\partial\theta_i} = \mathrm{Tr}\left[\left(\nabla_{\overline{\boldsymbol{\gamma}}}\,\ell\right)^\top\frac{\partial\overline{\boldsymbol{\gamma}}}{\partial\theta_i}\right] = \left(\nabla_{\overline{\boldsymbol{\gamma}}}\,\ell\right)^\top\frac{\partial\overline{\boldsymbol{\gamma}}}{\partial\theta_i}.$$

Therefore, it follows that the gradient of the log likelihood with respect to the mean parameters is

$$\nabla_{\boldsymbol{\theta}}\,\ell = \left(\nabla_{\overline{\boldsymbol{\gamma}}}\,\ell\right)^\top\frac{\partial\overline{\boldsymbol{\gamma}}}{\partial\boldsymbol{\theta}}.$$

The matrix derivative of the log likelihood with respect to $\mathbf{K}$ is

$$\begin{aligned}\frac{\partial\ell}{\partial\mathbf{K}} &= \frac{\partial}{\partial\mathbf{K}}\left[-\frac{1}{2}\log\det\mathbf{K}-\frac{1}{2}\left(\mathbf{U}^\top\hat{\boldsymbol{\gamma}}-\mathbf{A}\overline{\boldsymbol{\gamma}}\right)^\top\mathbf{K}^{-1}\left(\mathbf{U}^\top\hat{\boldsymbol{\gamma}}-\mathbf{A}\overline{\boldsymbol{\gamma}}\right)\right]\\ &= -\frac{1}{2}\mathbf{K}^{-1}+\frac{1}{2}\mathbf{K}^{-1}\left(\mathbf{U}^\top\hat{\boldsymbol{\gamma}}-\mathbf{A}\overline{\boldsymbol{\gamma}}\right)\left(\mathbf{U}^\top\hat{\boldsymbol{\gamma}}-\mathbf{A}\overline{\boldsymbol{\gamma}}\right)^\top\mathbf{K}^{-1}.\end{aligned}$$

The derivative of $\mathbf{K}$ with respect to a given covariance parameter is

$$\frac{\partial\mathbf{K}}{\partial\omega_i} = \mathbf{A}\frac{\partial\boldsymbol{\Sigma}}{\partial\omega_i}\mathbf{A}^\top.$$

Using the matrix chain rule, the derivative with respect to a given covariance parameter is

$$\frac{\partial\ell}{\partial\omega_i} = \mathrm{Tr}\left[\left(\frac{\partial\ell}{\partial\mathbf{K}}\right)^\top\frac{\partial\mathbf{K}}{\partial\omega_i}\right].$$

The Fisher information for multivariate normal distributions has a special form [5] such that the mean and covariance parameters do not share any information,

$$\boldsymbol{\mathcal{I}}\left(\boldsymbol{\theta},\boldsymbol{\omega}\right) = \begin{bmatrix}\boldsymbol{\mathcal{I}}_{\overline{\boldsymbol{\gamma}}}\left(\boldsymbol{\theta},\boldsymbol{\omega}\right) & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\mathcal{I}}_{\mathbf{K}}\left(\boldsymbol{\omega}\right)\end{bmatrix}.$$

The Fisher information of the mean parameter is

$$\left[\boldsymbol{\mathcal{I}}_{\overline{\boldsymbol{\gamma}}}\left(\boldsymbol{\theta},\boldsymbol{\omega}\right)\right]_{ij} = \left(\left.\frac{\partial\overline{\boldsymbol{\gamma}}}{\partial\theta_i}\right|_{\boldsymbol{\theta}}\right)^\top\mathbf{A}^\top\left[\mathbf{K}\left(\boldsymbol{\omega}\right)\right]^{-1}\mathbf{A}\left(\left.\frac{\partial\overline{\boldsymbol{\gamma}}}{\partial\theta_j}\right|_{\boldsymbol{\theta}}\right).$$

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

The Fisher information of the covariance parameters is

$$[\mathcal{I}_{\mathbf{K}}(\boldsymbol{\omega})]_{ij} = \frac{1}{2}\mathrm{Tr}\left[[\mathbf{K}(\boldsymbol{\omega})]^{-1}\left(\frac{\partial \mathbf{K}}{\partial \omega_i}\bigg|_{\boldsymbol{\omega}}\right)[\mathbf{K}(\boldsymbol{\omega})]^{-1}\left(\frac{\partial \mathbf{K}}{\partial \omega_j}\bigg|_{\boldsymbol{\omega}}\right)\right] .$$

Natural gradient ascent involves Newton-Raphson updates with the Fisher information matrix. A dampening parameter $0 < \alpha_t \leq 1$ is chosen using a backtracking line search [6] at each iteration to improve stability and serve as a stopping condition,

$$\hat{\boldsymbol{\theta}}_{t+1} = \hat{\boldsymbol{\theta}}_t + \alpha_t \left[\mathcal{I}_{\overline{\boldsymbol{\gamma}}}\left(\hat{\boldsymbol{\theta}}_t, \hat{\boldsymbol{\omega}}_t\right)\right]^{-1}\left[\nabla_{\boldsymbol{\theta}}\,\ell\,\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_t}\right]$$
$$\hat{\boldsymbol{\omega}}_{t+1} = \hat{\boldsymbol{\omega}}_t + \alpha_t \left[\mathcal{I}_{\mathbf{K}}(\hat{\boldsymbol{\omega}}_t)\right]^{-1}\left[\nabla_{\boldsymbol{\omega}}\,\ell\,\Big|_{\boldsymbol{\omega}=\hat{\boldsymbol{\omega}}_t}\right] .$$

In practice, we estimate the gradients and Fisher information matrix for each chromosome separately and sum them up because effect estimates are assumed to be independent across chromosomes. The derivatives $\frac{\partial \overline{\boldsymbol{\gamma}}}{\partial \theta_i}$ and $\frac{\partial \mathbf{K}}{\partial \omega_i}$ are computed using automatic differentiation [7, 8].

Uncertainty Estimation

Let $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\omega}}$ be the maximum likelihood estimates (MLEs) of $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$ respectively. Let $\psi(\boldsymbol{\theta}, \boldsymbol{\omega})$ be any continuous map from the parameters to a real number. ψ can itself be interpreted as a parameter of the model. In our model, ϕ is an example of such a parameter. A natural estimate for ψ is to map the MLEs,

$$\hat{\psi} = \psi\left(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\omega}}\right) .$$

By the properties of the MLE, the estimator converges to the following in distribution.

$$\hat{\boldsymbol{\theta}} \xrightarrow{\mathrm{d}} \mathcal{N}\left(\boldsymbol{\theta}, \left[\mathcal{I}_{\overline{\boldsymbol{\gamma}}}(\boldsymbol{\theta}, \boldsymbol{\omega})\right]^{-1}\right)$$
$$\hat{\boldsymbol{\omega}} \xrightarrow{\mathrm{d}} \mathcal{N}\left(\boldsymbol{\omega}, \left[\mathcal{I}_{\mathbf{K}}(\boldsymbol{\omega})\right]^{-1}\right)$$

By the delta method, the estimator for ψ converges to

$$\hat{\psi} \xrightarrow{\mathrm{d}} \mathcal{N}\left(\psi, \left[\nabla_{\boldsymbol{\theta}}\psi\Big|_{\boldsymbol{\theta},\boldsymbol{\omega}}\right]^{\top}\left[\mathcal{I}_{\overline{\boldsymbol{\gamma}}}(\boldsymbol{\theta}, \boldsymbol{\omega})\right]^{-1}\left[\nabla_{\boldsymbol{\theta}}\psi\Big|_{\boldsymbol{\theta},\boldsymbol{\omega}}\right] + \left[\nabla_{\boldsymbol{\omega}}\psi\Big|_{\boldsymbol{\theta},\boldsymbol{\omega}}\right]^{\top}\left[\mathcal{I}_{\mathbf{K}}(\boldsymbol{\omega})\right]^{-1}\left[\nabla_{\boldsymbol{\omega}}\psi\Big|_{\boldsymbol{\theta},\boldsymbol{\omega}}\right]\right) ,$$

which can be used to derive approximate confidence intervals. Confidence intervals for ϕ were estimated using this method.

Fixed-Effects Model

In addition to consistent estimators provided by maximum likelihood estimation, it is useful to have unbiased estimators of parameters as well. In the fixed-effects model, we assume that $\boldsymbol{\gamma}$ is fixed, implying

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

that the observations have the following likelihood:

$$\mathbf{U}^\top \hat{\boldsymbol{\gamma}} \mid \boldsymbol{\gamma} \sim \mathcal{N}\left(\mathbf{A}\boldsymbol{\gamma}, \boldsymbol{\Lambda}\right) .$$

For unbiased estimation, the projection matrix $\mathbf{A}^+\mathbf{A}$ is used frequently throughout our derivations. To see that it is a projection matrix, note that

$$\begin{aligned}
\mathbf{A}^+\mathbf{A} &= \hat{\mathbf{S}}^2\mathbf{U}\boldsymbol{\Lambda}^{-1}\mathbf{U}^\top\hat{\mathbf{S}}\hat{\mathbf{R}}\hat{\mathbf{S}}^{-1} \\
&= \hat{\mathbf{S}}^2\mathbf{U}\boldsymbol{\Lambda}^{-1}\mathbf{U}^\top\hat{\mathbf{S}}\hat{\mathbf{R}}\hat{\mathbf{S}}\hat{\mathbf{S}}^{-2} \\
&= \hat{\mathbf{S}}^2\mathbf{U}\boldsymbol{\Lambda}^{-1}\mathbf{U}^\top\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top\hat{\mathbf{S}}^{-2} \\
&= \hat{\mathbf{S}}^2\mathbf{U}\mathbf{U}^\top\hat{\mathbf{S}}^{-2} .
\end{aligned}$$

Then, it follows that

$$\begin{aligned}
\left(\mathbf{A}^+\mathbf{A}\right)\left(\mathbf{A}^+\mathbf{A}\right) &= \hat{\mathbf{S}}^2\mathbf{U}\mathbf{U}^\top\hat{\mathbf{S}}^{-2}\hat{\mathbf{S}}^2\mathbf{U}\mathbf{U}^\top\hat{\mathbf{S}}^{-2} \\
&= \hat{\mathbf{S}}^2\mathbf{U}\mathbf{U}^\top\mathbf{U}\mathbf{U}^\top\hat{\mathbf{S}}^{-2} \\
&= \hat{\mathbf{S}}^2\mathbf{U}\mathbf{U}^\top\hat{\mathbf{S}}^{-2} \\
&= \mathbf{A}^+\mathbf{A} .
\end{aligned}$$

This makes $\mathbf{A}^+\mathbf{A}$ an idempotent matrix and implies that it is a projection matrix.

Mean Effect

Suppose that the mean effect size is $\overline{\gamma}_1$ for the first variant class and $\overline{\gamma}_2$ for the second variant class. Then,

$$\begin{aligned}
\hat{\overline{\gamma}}_1 &= \frac{1}{M}\begin{bmatrix}\mathbf{1}\\ \mathbf{0}\end{bmatrix}^\top \mathbf{A}^+\mathbf{U}^\top\hat{\boldsymbol{\gamma}} \\
\hat{\overline{\gamma}}_2 &= \frac{1}{M}\begin{bmatrix}\mathbf{0}\\ \mathbf{1}\end{bmatrix}^\top \mathbf{A}^+\mathbf{U}^\top\hat{\boldsymbol{\gamma}}
\end{aligned}$$

are unbiased estimators for these mean effect sizes under certain conditions discussed below. Since the estimator is the sum of normal random variables, we can easily construct confidence intervals from the variance. A straightforward computation shows that the variances of the estimators are

$$\begin{aligned}
\mathbb{V}\left[\hat{\overline{\gamma}}_1\right] &\triangleq \frac{1}{M^2}\begin{bmatrix}\mathbf{1}\\ \mathbf{0}\end{bmatrix}^\top \mathbf{A}^+\boldsymbol{\Lambda}\left(\mathbf{A}^+\right)^\top\begin{bmatrix}\mathbf{1}\\ \mathbf{0}\end{bmatrix} \\
\mathbb{V}\left[\hat{\overline{\gamma}}_2\right] &\triangleq \frac{1}{M^2}\begin{bmatrix}\mathbf{0}\\ \mathbf{1}\end{bmatrix}^\top \mathbf{A}^+\boldsymbol{\Lambda}\left(\mathbf{A}^+\right)^\top\begin{bmatrix}\mathbf{0}\\ \mathbf{1}\end{bmatrix} .
\end{aligned}$$

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

These estimators are guaranteed to be unbiased if $\hat{\mathbf{S}}^{-2}\left[\mathbf{1}\ \mathbf{0}\right]^\top$ and $\hat{\mathbf{S}}^{-2}\left[\mathbf{0}\ \mathbf{1}\right]^\top$ are orthogonal to the null space of $\mathbf{U}^\top$.

To see that $\hat{\bar{\gamma}}_1$ and $\hat{\bar{\gamma}}_2$ are unbiased under these conditions, we will focus on the former without loss of generality. In expectation,

$$\begin{aligned}
\mathbb{E}\left[\begin{bmatrix}\mathbf{1}\\ \mathbf{0}\end{bmatrix}^\top \mathbf{A}^+\mathbf{U}^\top\hat{\boldsymbol{\gamma}}\right] &= \begin{bmatrix}\mathbf{1}\\ \mathbf{0}\end{bmatrix}^\top \mathbf{A}^+\mathbb{E}\left[\mathbf{U}^\top\hat{\boldsymbol{\gamma}}\right] \\
&= \begin{bmatrix}\mathbf{1}\\ \mathbf{0}\end{bmatrix}^\top \mathbf{A}^+\mathbf{A}\boldsymbol{\gamma} \\
&= \begin{bmatrix}\mathbf{1}\\ \mathbf{0}\end{bmatrix}^\top \mathbf{A}^+\mathbf{A}\left(\begin{bmatrix}\boldsymbol{\gamma}_1\\ \mathbf{0}\end{bmatrix} + \begin{bmatrix}\mathbf{0}\\ \boldsymbol{\gamma}_2\end{bmatrix}\right) \\
&= \begin{bmatrix}\mathbf{1}\\ \mathbf{0}\end{bmatrix}^\top \mathbf{A}^+\mathbf{A}\begin{bmatrix}\boldsymbol{\gamma}_1\\ \mathbf{0}\end{bmatrix} + \begin{bmatrix}\mathbf{1}\\ \mathbf{0}\end{bmatrix}^\top \mathbf{A}^+\mathbf{A}\begin{bmatrix}\mathbf{0}\\ \boldsymbol{\gamma}_2\end{bmatrix} \\
&= \begin{bmatrix}\mathbf{1}\\ \mathbf{0}\end{bmatrix}^\top \mathbf{A}^+\mathbf{A}\begin{bmatrix}\boldsymbol{\gamma}_1\\ \mathbf{0}\end{bmatrix}.
\end{aligned}$$

The last equality holds because $\mathbf{A}^+\mathbf{A}$ is block diagonal. Note here that $\mathbf{A}^+\mathbf{A}$ is a projection matrix and has the form $\mathbf{A}^+\mathbf{A} = \hat{\mathbf{S}}^2\mathbf{U}\mathbf{U}^\top\hat{\mathbf{S}}^{-2}$. Therefore, for our estimator to be unbiased, we require that

$$\begin{bmatrix}\mathbf{1}\\ \mathbf{0}\end{bmatrix}^\top \mathbf{A}^+\mathbf{A} = \begin{bmatrix}\mathbf{1}\\ \mathbf{0}\end{bmatrix}^\top,$$

which is true if $\hat{\mathbf{S}}^{-2}\left[\mathbf{1}\ \mathbf{0}\right]^\top$ is orthogonal to the null space of $\mathbf{U}^\top$. A similar derivation demonstrates that for $\hat{\bar{\gamma}}_2$ to be unbiased, $\hat{\mathbf{S}}^{-2}\left[\mathbf{0}\ \mathbf{1}\right]^\top$ must be orthogonal to the null space of $\mathbf{U}^\top$. We tested this numerically with the eigendecomposition involving our specific LD matrix.

Mean Squared Effect

The mean squared effect is a useful measure of the amount of burden signal present for each trait. In this paper, we estimate the squared effect for the LoF burden effects. The LoF burden genotypes are assumed to not be in LD, meaning that the likelihood for the jth gene is

$$\hat{\gamma}_{1j} \mid \gamma_{1j} \sim \mathcal{N}\left(\gamma_{1j}, s_{1j}^2\right).$$

Since the second moment of a univariate normal random variable is

$$\mathbb{E}\left[\hat{\gamma}_{1j}^2\right] = \gamma_{1j}^2 + s_{1j}^2,$$

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

we note that

$$\widehat{\overline{\gamma}_1^2} \triangleq \frac{1}{M}\sum_{j=1}^{M} \hat{\gamma}_{1j}^2 - s_{1j}^2$$

is an unbiased estimate of the squared burden effect size.

We invoke the Central Limit Theorem, noting that each element in the sum is an independent and unbiased estimate of the quantity of interest, to derive approximate confidence intervals. We use an empirical estimate of the standard error to derive the approximate confidence intervals.

Random-Effects Model

A random-effects model is required to develop estimators for the covariance components. We use an approach inspired by linkage disequilibrium score regression (LDSC) [9], and similar to that outlined in [3]. We assume that $\boldsymbol{\gamma}$ is random, with finite first and second moments. We also assume that the latent effect sizes are independent given $\hat{\mathbf{S}}$ and $\hat{\mathbf{R}}$. We make the following minimal assumptions about first and second moments of the joint distribution of the latent effect sizes:

$$\mathbb{E}\left[\boldsymbol{\gamma} \mid \hat{\mathbf{S}}, \hat{\mathbf{R}}\right] = \overline{\boldsymbol{\gamma}}$$
$$\mathbb{V}\left[\boldsymbol{\gamma} \mid \hat{\mathbf{S}}, \hat{\mathbf{R}}\right] = \boldsymbol{\Sigma}\,.$$

The Z scores from the burden regression can be defined as

$$\mathbf{z} = \hat{\mathbf{S}}^{-1}\hat{\boldsymbol{\gamma}}.$$

By the properties of the normal distribution, the Z scores are distributed as

$$\mathbf{z} \mid \hat{\mathbf{S}}, \hat{\mathbf{R}}, \boldsymbol{\gamma} \sim \mathcal{N}\left(\hat{\mathbf{R}}\hat{\mathbf{S}}^{-1}\boldsymbol{\gamma}, \hat{\mathbf{R}}\right)\,,$$

and Z scores for the two variant classes are independent,

$$\mathbf{z}_1 \mid \hat{\mathbf{S}}, \hat{\mathbf{R}}, \boldsymbol{\gamma} \sim \mathcal{N}\left(\hat{\mathbf{R}}_1\hat{\mathbf{S}}_1^{-1}\boldsymbol{\gamma}_1, \hat{\mathbf{R}}_1\right)$$
$$\mathbf{z}_2 \mid \hat{\mathbf{S}}, \hat{\mathbf{R}}, \boldsymbol{\gamma} \sim \mathcal{N}\left(\hat{\mathbf{R}}_2\hat{\mathbf{S}}_2^{-1}\boldsymbol{\gamma}_2, \hat{\mathbf{R}}_2\right)\,.$$

Without loss of generality for the second variant class, we focus on the jth gene of the first variant class. Let $\hat{r}_{1ji}$ be the jith element of $\hat{\mathbf{R}}_1$. The marginal distribution of the Z score for a given gene is then

$$z_{1j} \mid \hat{\mathbf{S}}, \hat{\mathbf{R}}, \boldsymbol{\gamma} \sim \mathcal{N}\left(\sum_{i=1}^{M} \frac{\hat{r}_{1ji}}{\hat{s}_{1i}}\gamma_{1i}, 1\right)\,.$$

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

We define the scaled LD score for the jth gene as

$$\ell_{1j} \triangleq \sum_{i=1}^{M} \frac{\hat{r}_{1ji}^2}{\hat{s}_{1i}^2} \,.$$

It is also useful to define

$$l_{1j} \triangleq \sum_{i=1}^{M} \frac{\hat{r}_{1ji}}{\hat{s}_{1i}} \,.$$

The marginal distribution of the Z scores between the two variant classes is

$$\mathbf{z}_{\cdot j} \mid \hat{\mathbf{S}}, \hat{\mathbf{R}}, \boldsymbol{\gamma} \sim \mathcal{N}\left(\begin{bmatrix} \sum_{i=1}^{M} \hat{r}_{1ji} \hat{s}_{1j}^{-1} \gamma_{1i} \\ \sum_{i=1}^{M} \hat{r}_{2ji} \hat{s}_{2j}^{-1} \gamma_{2i} \end{bmatrix}, \mathbf{I} \right) \,.$$

We define the cross-variant LD score as

$$\ell_j \triangleq \sum_{i=1}^{M} \frac{\hat{r}_{1ji} \hat{r}_{2ji}}{\hat{s}_{1j} \hat{s}_{2j}} \,.$$

It is also useful to define

$$l_j \triangleq \left[\sum_{i=1}^{M} \frac{\hat{r}_{1ji}}{\hat{s}_{1j}} \right] \left[\sum_{i=1}^{M} \frac{\hat{r}_{2ji}}{\hat{s}_{2j}} \right] \,.$$

Covariance Components

Similar to LDSC, we now proceed by evaluating the expected squared Z score to develop estimators for σ_{11}^2 and σ_{22}^2. We integrate over the latent effect size, $\boldsymbol{\gamma}$, using the law of total expectation and the law of total variance. In the following, the conditioning on $\hat{\mathbf{R}}_1$, $\hat{\mathbf{S}}_1$, and ℓ_{1j} is implicit to simplify notation. The expected squared Z score is

$$\begin{aligned} \mathbb{E}\left[z_{1j}^2\right] &= \mathbb{V}\left[z_{1j}\right] + \left(\mathbb{E}\left[z_{1j}\right]\right)^2 \\ &= \mathbb{V}\left[\mathbb{E}\left[z_{1j} \mid \boldsymbol{\gamma}\right]\right] + \mathbb{E}\left[\mathbb{V}\left[z_{1j} \mid \boldsymbol{\gamma}\right]\right] + \left(\mathbb{E}\left[\mathbb{E}\left[z_{1j} \mid \boldsymbol{\gamma}\right]\right]\right)^2 \,. \end{aligned}$$

The first term introduces the LD score into the expectation,

$$\begin{aligned} \mathbb{V}\left[\mathbb{E}\left[z_{1j} \mid \boldsymbol{\gamma}\right]\right] &= \mathbb{V}\left[\sum_{i=1}^{M} \frac{\hat{r}_{1ji}}{\hat{s}_{1i}} \gamma_{1i}\right] \\ &= \sum_{i=1}^{M} \frac{\hat{r}_{1ji}^2}{\hat{s}_{1i}^2} \mathbb{V}\left[\gamma_{1i}\right] \\ &= \sigma_{11}^2 \ell_{1j} \,. \end{aligned}$$

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

The second term introduces a constant,

$$\mathbb{E}\left[\mathbb{V}\left[z_{1j} \mid \boldsymbol{\gamma}\right]\right] = 1\,.$$

The last term is a bias introduced by the uncentered prior,

$$\begin{aligned}\mathbb{E}\left[\mathbb{E}\left[z_{j1} \mid \boldsymbol{\gamma}\right]\right] &= \mathbb{E}\left[\sum_{i=1}^{M} \frac{\hat{r}_{1ji}}{\hat{s}_{1i}} \gamma_{1i}\right] \\ &= \sum_{i=1}^{M} \frac{\hat{r}_{1ji}}{\hat{s}_{1i}} \mathbb{E}\left[\gamma_{1i}\right] \\ &= \overline{\gamma}_1 l_{1j}\,.\end{aligned}$$

Taken together, the expected squared Z score is

$$\mathbb{E}\left[z_{1j}^2\right] = \sigma_{11}^2 \ell_{1j} + \overline{\gamma}_1^2 l_{1j}^2 + 1\,.$$

Given the relationship of the expected squared Z score with σ_{11}^2, we propose the following method-of-moments (MoM) estimator. A similar derivation follows for σ_{22}^2. We use $\hat{\overline{\gamma}}_1$, which is unbiased for $\overline{\gamma}_1$ under the random effects model as well. Our estimators are

$$\begin{aligned}\hat{\sigma}_{11}^2 &\triangleq \frac{1}{M} \sum_{j=1}^{M} \frac{z_{1j}^2 - \left(\hat{\overline{\gamma}}_1 l_{1j}\right)^2 - 1}{\ell_{1j}} \\ \hat{\sigma}_{22}^2 &\triangleq \frac{1}{M} \sum_{j=1}^{M} \frac{z_{2j}^2 - \left(\hat{\overline{\gamma}}_2 l_{2j}\right)^2 - 1}{\ell_{2j}}\,.\end{aligned}$$

We can use the expected product of Z scores to build an estimator for the covariance σ_{12}. We integrate over the latent effect size $\boldsymbol{\gamma}$ using the law of total expectation and the law of total covariance. In the following, the conditioning on $\hat{\mathbf{R}}$, $\hat{\mathbf{S}}$, and ℓ_j is implicit to simplify notation. The expected product of Z scores is

$$\begin{aligned}\mathbb{E}\left[z_{1j} z_{2j}\right] &= \mathbb{V}\left[z_{1j}, z_{2j}\right] + \mathbb{E}\left[z_{1j}\right] \mathbb{E}\left[z_{2j}\right] \\ &= \mathbb{V}\left[\mathbb{E}\left[z_{1j} \mid \boldsymbol{\gamma}\right], \mathbb{E}\left[z_{2j} \mid \boldsymbol{\gamma}\right]\right] + \mathbb{E}\left[\mathbb{V}\left[z_{1j}, z_{2j} \mid \boldsymbol{\gamma}\right]\right] + \mathbb{E}\left[\mathbb{E}\left[z_{1j} \mid \boldsymbol{\gamma}\right]\right] \mathbb{E}\left[\mathbb{E}\left[z_{2j} \mid \boldsymbol{\gamma}\right]\right]\,.\end{aligned}$$

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

The first term introduces the cross-variant LD score into the expectation,

$$\begin{aligned}\mathbb{V}\left[\mathbb{E}\left[z_{1j} \mid \boldsymbol{\gamma}\right], \mathbb{E}\left[z_{2j} \mid \boldsymbol{\gamma}\right]\right] &= \mathbb{V}\left[\sum_{i=1}^{M} \frac{\hat{r}_{1ji}}{\hat{s}_{1i}}\gamma_{1i}, \sum_{k=1}^{M} \frac{\hat{r}_{2jk}}{\hat{s}_{2k}}\gamma_{2k}\right] \\ &= \sum_{i=1}^{M}\sum_{k=1}^{M} \frac{\hat{r}_{1ji}\hat{r}_{2jk}}{\hat{s}_{1i}\hat{s}_{2k}}\mathbb{V}\left[\gamma_{1i}, \gamma_{2k}\right] \\ &= \sum_{i=1}^{M} \frac{\hat{r}_{1ji}\hat{r}_{2ji}}{\hat{s}_{1i}\hat{s}_{2i}}\mathbb{V}\left[\gamma_{1i}, \gamma_{2i}\right] \\ &= \sigma_{12}\ell_j\,.\end{aligned}$$

The second term does not contribute to the conditional expectation because the covariance between observed effect sizes is zero given the latent effect size:

$$\mathbb{E}\left[\mathbb{V}\left[z_{1j}, z_{2j} \mid \boldsymbol{\gamma}\right]\right] = 0\,.$$

The last term is bias introduced by the uncentered prior,

$$\begin{aligned}\mathbb{E}\left[\mathbb{E}\left[z_{1j} \mid \boldsymbol{\gamma}\right]\right]\mathbb{E}\left[\mathbb{E}\left[z_{2j} \mid \boldsymbol{\gamma}\right]\right] &= \mathbb{E}\left[\sum_{i=1}^{M} \frac{\hat{r}_{1ji}}{\hat{s}_{1i}}\gamma_{1i}\right]\mathbb{E}\left[\sum_{k=1}^{M} \frac{\hat{r}_{2jk}}{\hat{s}_{2k}}\gamma_{2k}\right] \\ &= \left[\sum_{i=1}^{M} \frac{\hat{r}_{1ji}}{\hat{s}_{1i}}\mathbb{E}\left[\gamma_{1i}\right]\right]\left[\sum_{k=1}^{M} \frac{\hat{r}_{2jk}}{\hat{s}_{2k}}\mathbb{E}\left[\gamma_{2k}\right]\right] \\ &= \overline{\gamma}_1\overline{\gamma}_2 l_{1j}l_{2j}\,.\end{aligned}$$

Taken together, the expected product of Z scores is

$$\mathbb{E}\left[z_{1j}z_{2j}\right] = \sigma_{12}\ell_j + \overline{\gamma}_1\overline{\gamma}_2 l_{1j}l_{2j}\,.$$

This suggests the use of the following unbiased estimator for σ_{12},

$$\hat{\sigma}_{12} \triangleq \frac{1}{M}\sum_{i=1}^{M} \frac{z_{1j}z_{2j} - \hat{\overline{\gamma}}_1\hat{\overline{\gamma}}_2 l_{1j}l_{2j}}{\ell_j}\,.$$

Monotonicity

These estimators of individual parameters in the model allow us to define the MoM estimator for monotonicity as

$$\hat{\phi} \triangleq -\frac{\hat{\overline{\gamma}}_1\hat{\overline{\gamma}}_2 + \hat{\sigma}_{12}}{\sqrt{\left(\hat{\overline{\gamma}}_1^2 + \hat{\sigma}_{11}^2\right)\left(\hat{\overline{\gamma}}_2^2 + \hat{\sigma}_{22}^2\right)}}\,.$$

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

MoM estimators are not unbiased, but they are consistent. Since the estimator represents a function of various estimators, we used the bootstrap with 1000 iterations to estimate the standard error of the estimate. If the estimate was undefined due to a negative square root, we set $\hat{\phi}$ to zero.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

D Simulations

We used the UK Biobank (UKB) LoF and duplication burden genotypes for simulation. These simulations are not from the models we used to derive the various estimators. Thus, these tests are fair evaluations of our estimators and also demonstrate the validity of using the RSS likelihood to model burden summary statistics. Some of our estimators assume fixed effect sizes, while others assume random effect sizes. We performed two separate suites of simulations for these estimators.

The fixed-effect simulations assumed that each gene had some fixed effect $\overline{\gamma}_1$ via LoF variants and $\overline{\gamma}_2$ via duplication variants. We also performed fixed-effect simulations where each gene had a unique effect, with a genome-wide average effect of $\overline{\gamma}_1$ and $\overline{\gamma}_2$ respectively. The random-effect simulations assumed that the effect size for each gene was generated from

$$\boldsymbol{\gamma}_{\cdot j} \sim \mathcal{N}\left(\begin{bmatrix}\overline{\gamma}_1 \\ \overline{\gamma}_2\end{bmatrix}, \begin{bmatrix}\sigma_{11}^2 & \sigma_{12} \\ \sigma_{12} & \sigma_{22}^2\end{bmatrix}\right).$$

We simulated effect sizes for genes that had both LoF and duplication burden genotypes, corresponding to passing our genotyping filters described in our methods. Once effect sizes were assigned to each gene, we used PLINK [10] to score each sample based on the effect sizes. This provided us with genetic values for each individual under the simulation. Next, we added noise using draws from a normal distribution such that the expected variance of the phenotype would be one.

We used PLINK to run a single-variant association scan with the burden genotypes. We included 15 genotyping principal components (PCs), 20 rare variant PCs, and genotyping batch as covariates. Since we simulated from the actual genotypes, we had to include these covariates to control for confounding from population structure. Each parameter value was tested with 100 simulations. When perturbing one parameter, the other parameters were held constant at $\overline{\gamma}_1 = 0$, $\overline{\gamma}_2 = 0$, $\sigma_{11}^2 = 0.01$, $\sigma_{12} = 0$, and $\sigma_{22}^2 = 0.01$. The default variance values are based on the order of magnitude of burden heritability estimates from prior work [11].

Unbiased Average Burden Effect Estimators

Fixed-effect simulations were used to test $\hat{\overline{\gamma}}_1$ and $\hat{\overline{\gamma}}_2$. These estimators were convincingly unbiased for the various realistic values that we simulated (Figure D.1). We also developed estimators for the variance of these fixed-effect average burden effect estimators. Confidence intervals built using these variance estimates are well-calibrated (Figures D.2 and D.3). We also tested the estimators under a regime where each gene had a unique effect size, with relatively unbiased behavior and good calibration (Figures D.4, D.5, and D.6).

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

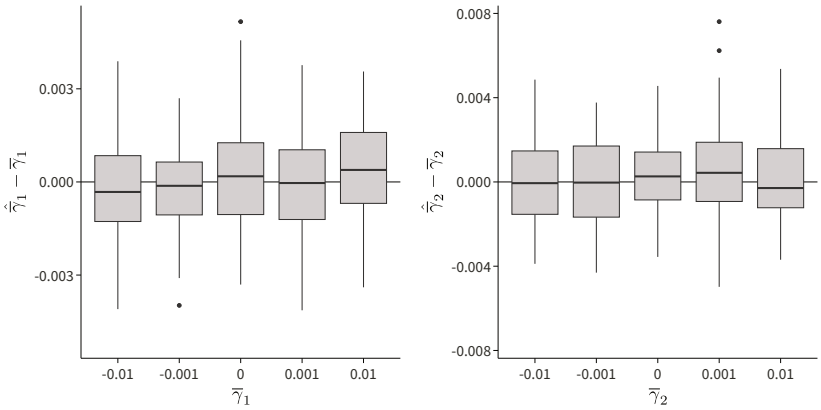

FIGURE D.1. (Left) The difference in the estimated value $\hat{\overline{\gamma}}_1$ and the true value $\overline{\gamma}_1$ for various tested values are shown for 100 simulations. (Right) The difference in the estimated value $\hat{\overline{\gamma}}_2$ and the true value $\overline{\gamma}_2$ for various tested values are shown for 100 simulations.

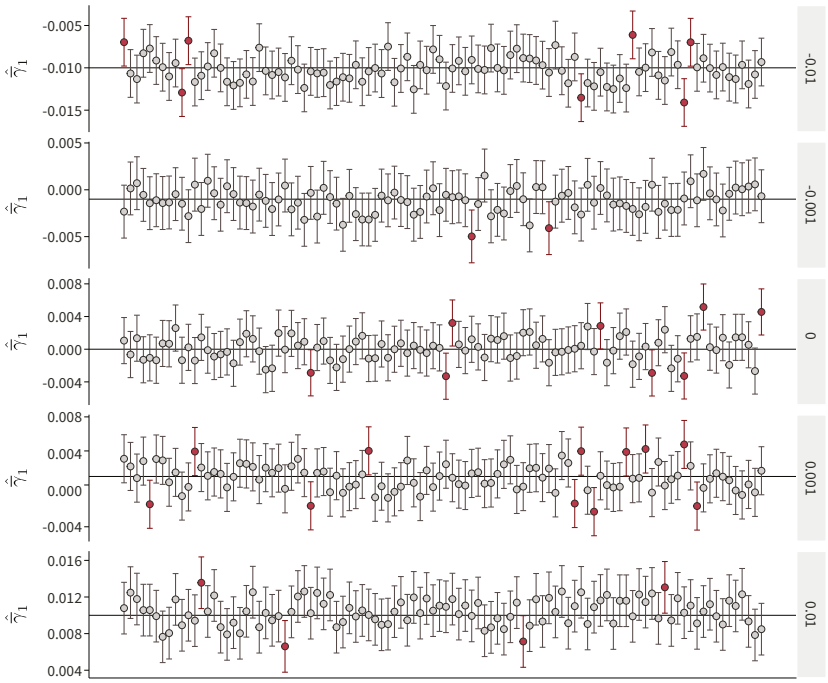

FIGURE D.2. The plot shows 95% confidence intervals from 100 simulations for various values of $\overline{\gamma}_1$. Confidence intervals in red represent realizations that do not cover the true value.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

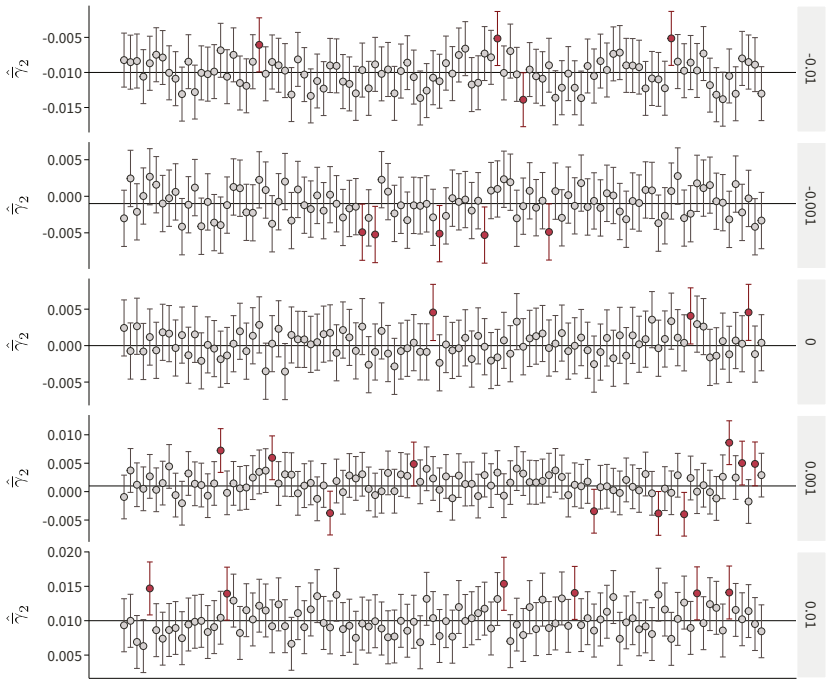

FIGURE D.3. The plot shows 95% confidence intervals from 100 simulations for various values of $\overline{\gamma}_2$. Confidence intervals in red represent realizations that do not cover the true value.

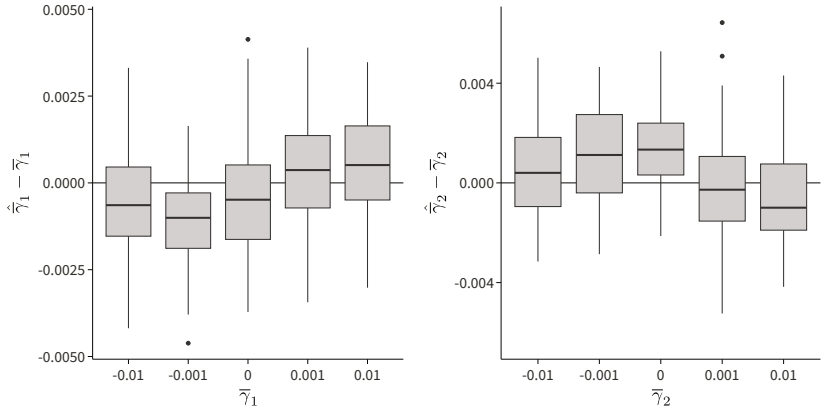

FIGURE D.4. (Left) The difference in the estimated value $\hat{\overline{\gamma}}_1$ and the true value $\overline{\gamma}_1$ for various tested values are shown for 100 simulations. (Right) The difference in the estimated value $\hat{\overline{\gamma}}_2$ and the true value $\overline{\gamma}_2$ for various tested values are shown for 100 simulations.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

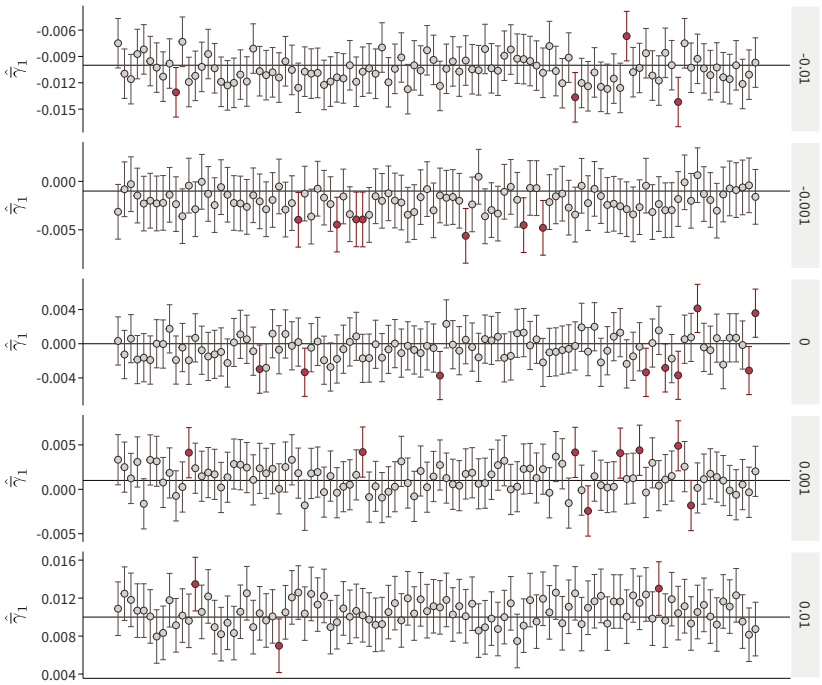

FIGURE D.5. The plot shows 95% confidence intervals from 100 simulations for various values of $\overline{\gamma}_1$. Confidence intervals in red represent realizations that do not cover the true value.

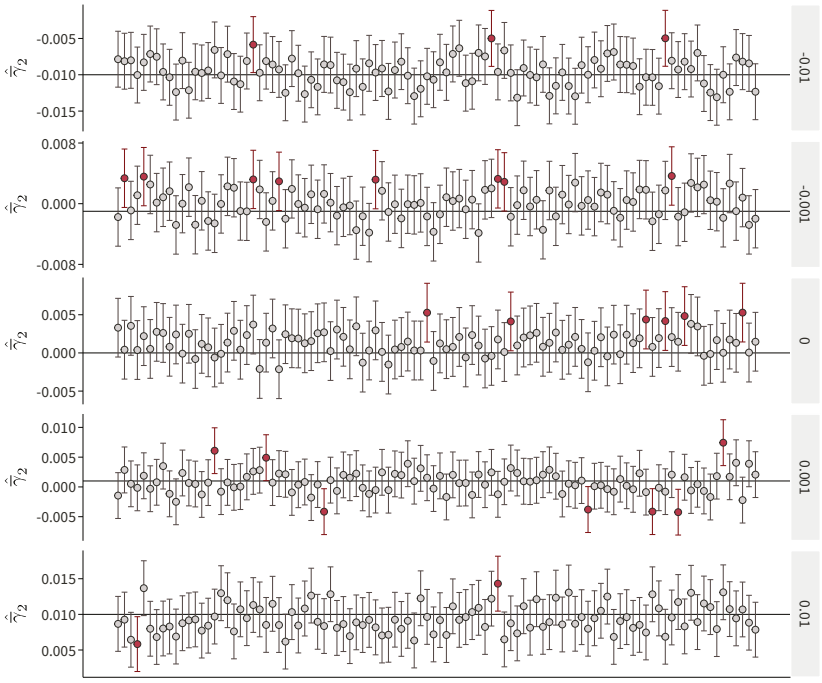

FIGURE D.6. The plot shows 95% confidence intervals from 100 simulations for various values of $\overline{\gamma}_2$. Confidence intervals in red represent realizations that do not cover the true value.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

Unbiased Average Squared Burden Effect Estimator

We used the fixed-effect simulations with different gene-level effects to test $\widehat{\overline{\gamma}_1^2}$. The estimator was relatively unbiased for various realistic values that we simulated (Figure D.7). The approximate confidence intervals using the Central Limit Theorem are well-calibrated (Figure D.8).

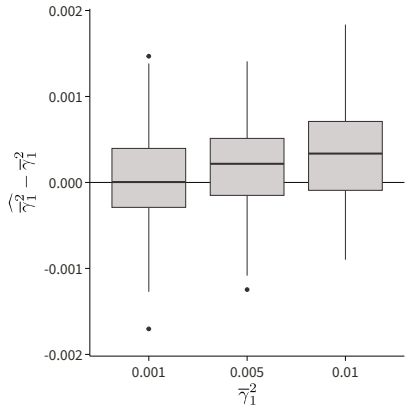

Figure D.7. The difference in the estimated value $\widehat{\overline{\gamma}_1^2}$ and the true value $\overline{\gamma}_1^2$ for various tested values are shown for 100 simulations.

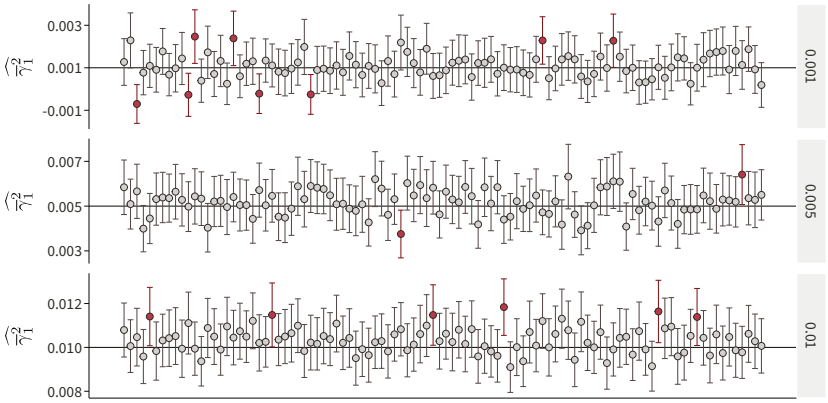

Figure D.8. The plot shows 95% confidence intervals from 100 simulations for various values of $\overline{\gamma}_1^2$. Confidence intervals in red represent realizations that do not cover the true value.

Method-of-Moments Monotonicity Estimator

Our covariance component estimators were close to unbiased (Figure D.9). We noted a slight upward bias in the LoF burden effect variance estimator $\hat{\sigma}_{22}^2$. However, the magnitude of this bias was only around 5% of the true parameter value. We believe that this represents residual confounding that is not corrected for using the genotyping and rare variant PCs used in our simulations.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

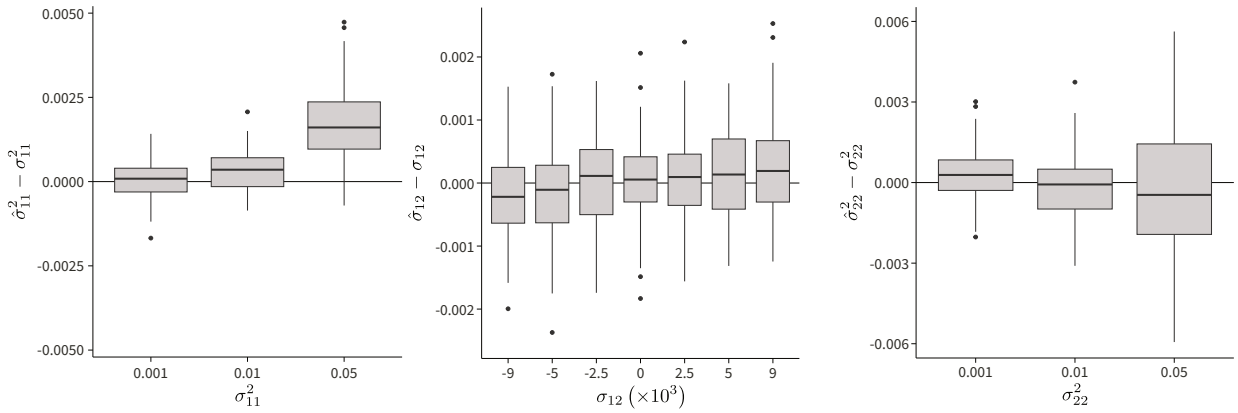

FIGURE D.9. (Left) The difference in the estimated value $\hat{\sigma}^2_{11}$ and the true value σ^2_{11} for various tested values are shown for 100 simulations. (Center) The difference in the estimated value $\hat{\sigma}_{12}$ and the true value σ_{12} for various tested values are shown for 100 simulations. (Right) The difference in the estimated value $\hat{\sigma}^2_{22}$ and the true value σ^2_{22} for various tested values are shown for 100 simulations.

Our MoM estimator for monotonicity was surprisingly close to unbiased, even for extreme deviations from zero (Figure D.10). The confidence intervals estimated using the bootstrap were also relatively well-calibrated (Figure D.11).

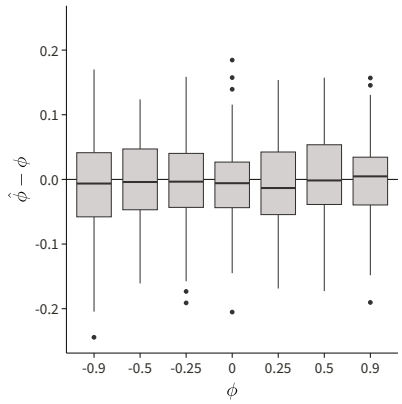

FIGURE D.10. The difference in the estimated value $\hat{\phi}$ and the true value ϕ for various tested values are shown for 100 simulations.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

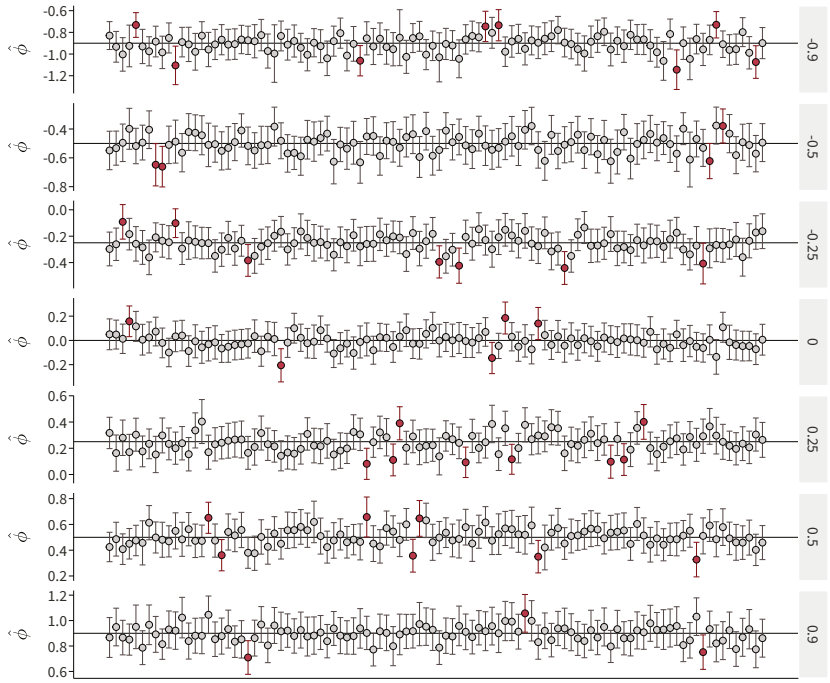

FIGURE D.11. The plot shows 95% confidence intervals from 100 simulations for various values of ϕ. Confidence intervals in red represent realizations that do not cover the true value.

MAXIMUM LIKELIHOOD AVERAGE BURDEN EFFECT ESTIMATORS

For the maximum likelihood estimation approach, we used random-effect simulations to test $\hat{\overline{\gamma}}_1$ and $\hat{\overline{\gamma}}_2$. These estimators were also convincingly unbiased, similar to our unbiased estimators (Figure D.12). The approximate confidence intervals are well-calibrated (Figures D.13 and D.14).

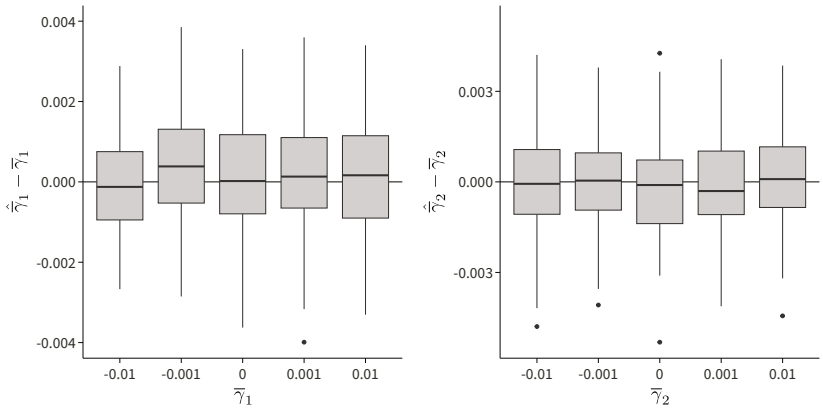

FIGURE D.12. (Left) The difference in the estimated value $\hat{\overline{\gamma}}_1$ and the true value $\overline{\gamma}_1$ for various tested values are shown for 100 simulations. (Right) The difference in the estimated value $\hat{\overline{\gamma}}_2$ and the true value $\overline{\gamma}_2$ for various tested values are shown for 100 simulations.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

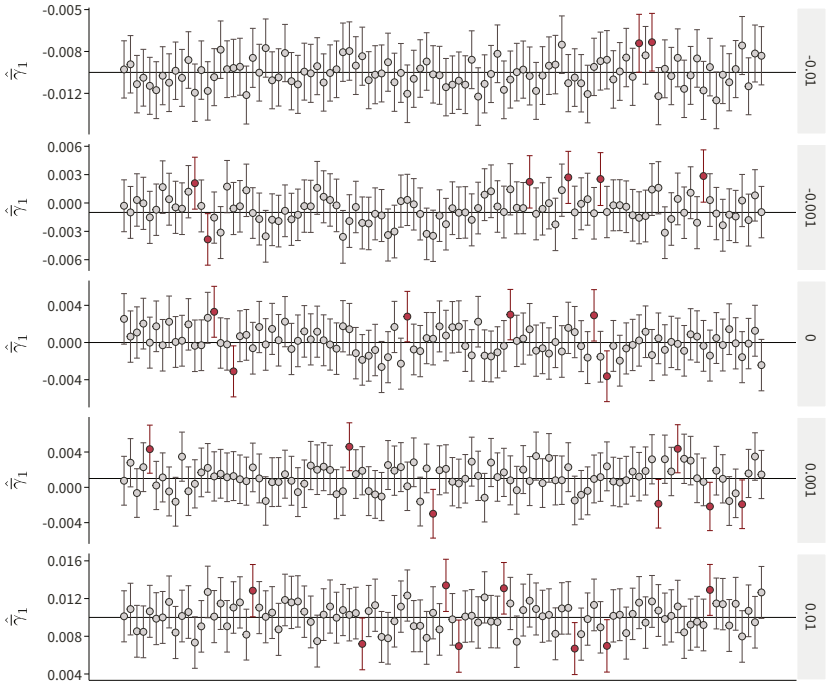

FIGURE D.13. The plot shows 95% confidence intervals from 100 simulations for various values of $\overline{\gamma}_1$. Confidence intervals in red represent realizations that do not cover the true value.

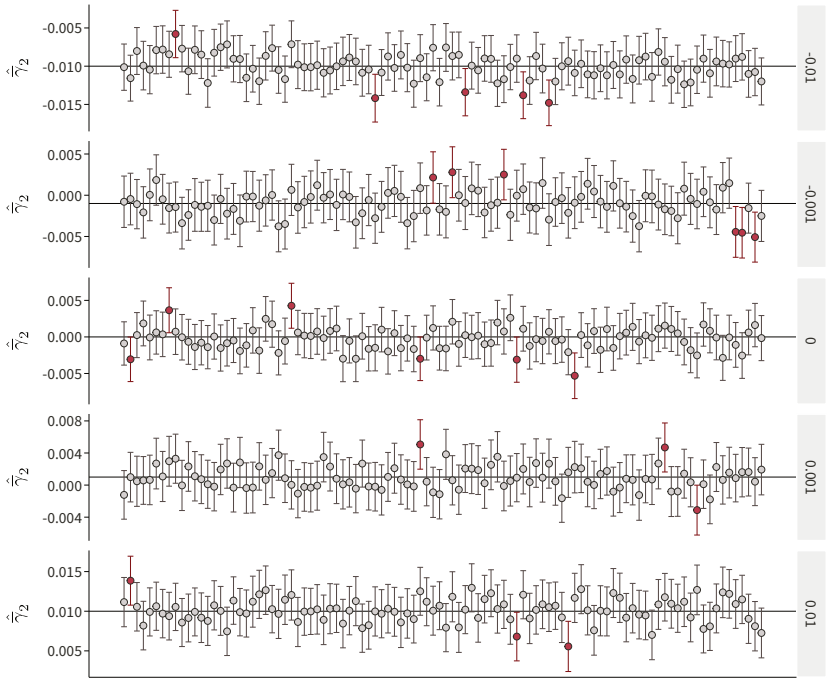

FIGURE D.14. The plot shows 95% confidence intervals from 100 simulations for various values of $\overline{\gamma}_2$. Confidence intervals in red represent realizations that do not cover the true value.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

Maximum Likelihood Monotonicity Estimator

Our covariance component estimators from the maximum likelihood approach had a similar performance to the unbiased estimators (Figure D.15). We noted a slight upward bias in the LoF burden effect variance estimator, similar to the unbiased estimator. The bias was again only around 5% of the true parameter value.

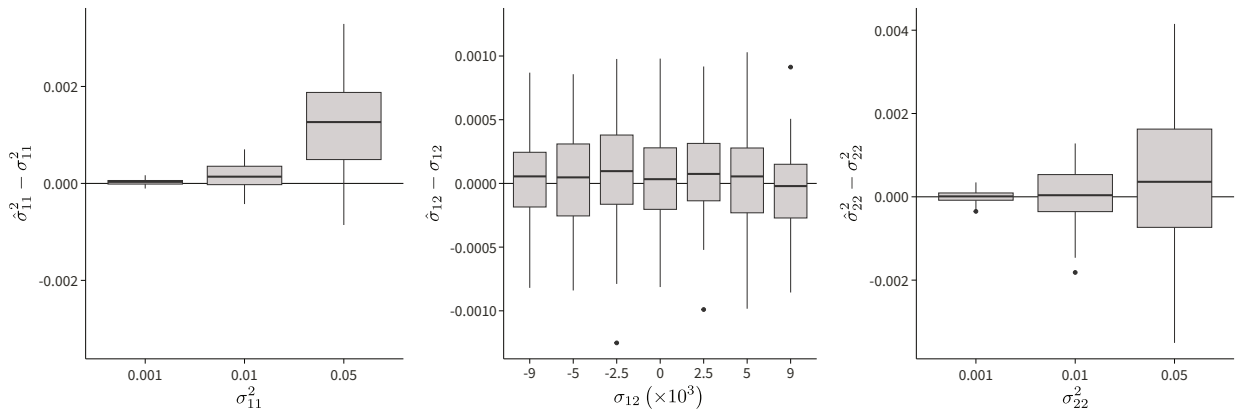

Figure D.15. (Left) The difference in the estimated value $\hat{\sigma}^2_{11}$ and the true value σ^2_{11} for various tested values are shown for 100 simulations. (Center) The difference in the estimated value $\hat{\sigma}_{12}$ and the true value σ_{12} for various tested values are shown for 100 simulations. (Right) The difference in the estimated value $\hat{\sigma}^2_{22}$ and the true value σ^2_{22} for various tested values are shown for 100 simulations.

The maximum likelihood estimate for monotonicity was slightly biased towards zero for very large values of ϕ (Figure D.16). The magnitude of bias was small compared to the true value, and the approximate confidence intervals using the delta method remained well-calibrated (Figure D.17). The variance of the estimator was approximately half that of the MoM estimator (Figure D.10).

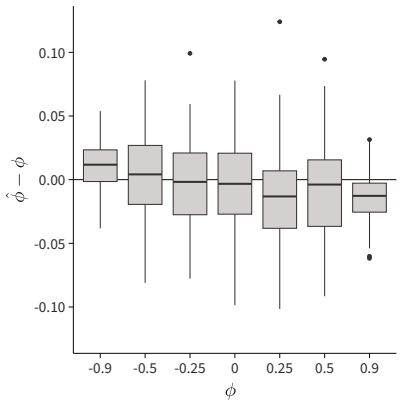

Figure D.16. The difference in the estimated value $\hat{\phi}$ and the true value ϕ for various tested values are shown for 100 simulations.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

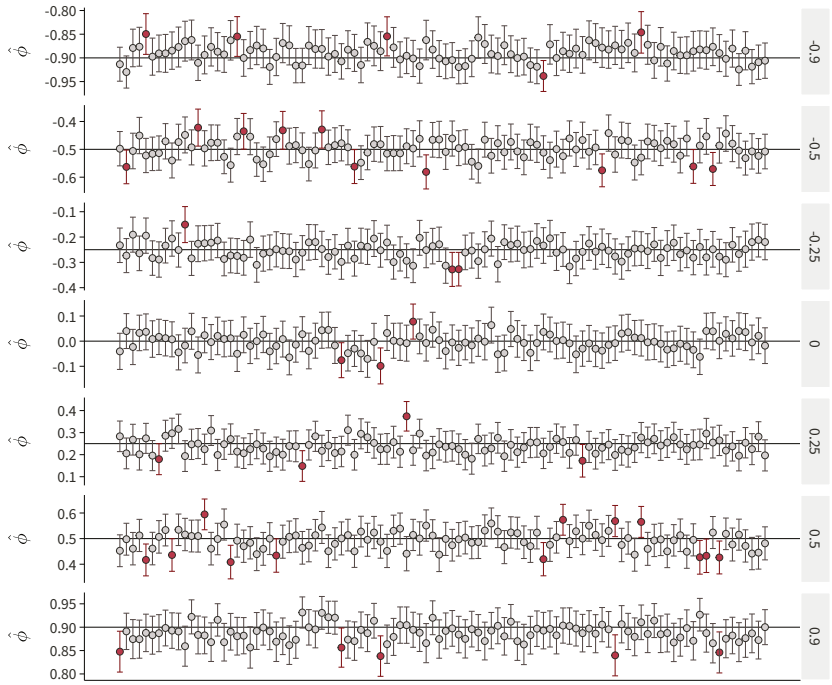

FIGURE D.17. The plot shows 95% confidence intervals from 100 simulations for various values of ϕ. Confidence intervals in red represent realizations that do not cover the true value.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

E Buffering Model

In the buffering model, we are interested in evaluating if GDRCs are systematically buffered against one trait direction. This could occur if non-monotone GDRCs are preferentially present in one direction over the other. This can also occur if monotone GDRCs achieve larger values for the trait in one direction compared to the other.

We use the multivariate adaptive shrinkage (MASH) model [12] as a flexible prior for our latent effect sizes. The MASH prior consists of a mixture of multivariate normal distributions. Suppose $\{\mathbf{V}_k\}_{k=1}^K$ represents a fixed set of covariance matrices in $\mathbb{R}^{2\times 2}$ for K mixture components. We use bivariate covariance matrices over a grid of variances and correlation values. Let $\boldsymbol{\pi} \in \Delta_{K-1}$ represent the mixture weights. Then, the gene-level prior is

$$\boldsymbol{\gamma}_{\cdot j} \mid \boldsymbol{\pi} \sim \sum_{k=1}^{K} \pi_k \mathcal{N}\left(\mathbf{0}, \mathbf{V}_k\right) .$$

The prior over the entire set of M genes is

$$\boldsymbol{\gamma} \mid \boldsymbol{\pi} \sim \prod_{j=1}^{M} \sum_{k=1}^{K} \pi_k \mathcal{N}\left(\boldsymbol{\gamma}_{\cdot j}\,; \mathbf{0}, \mathbf{V}_k\right) .$$

The likelihood model remains the same as before.

$$\mathbf{U}^\top \hat{\boldsymbol{\gamma}} \mid \boldsymbol{\gamma} \sim \mathcal{N}\left(\mathbf{U}^\top \hat{\mathbf{S}} \hat{\mathbf{R}} \hat{\mathbf{S}}^{-1} \boldsymbol{\gamma}, \boldsymbol{\Lambda}\right)$$

The joint likelihood is

$$\begin{aligned} p\left(\mathbf{U}^\top \hat{\boldsymbol{\gamma}}, \boldsymbol{\gamma} \mid \boldsymbol{\pi}\right) &= p\left(\mathbf{U}^\top \hat{\boldsymbol{\gamma}} \mid \boldsymbol{\gamma}\right) p\left(\boldsymbol{\gamma} \mid \boldsymbol{\pi}\right) \\ &= \mathcal{N}\left(\mathbf{U}^\top \hat{\boldsymbol{\gamma}}\,; \mathbf{U}^\top \hat{\mathbf{S}} \hat{\mathbf{R}} \hat{\mathbf{S}}^{-1} \boldsymbol{\gamma}, \boldsymbol{\Lambda}\right) \prod_{j=1}^{M} \sum_{k=1}^{K} \pi_k \mathcal{N}\left(\boldsymbol{\gamma}_{\cdot j}\,; \mathbf{0}, \mathbf{V}_k\right) . \end{aligned}$$

Inference using Empirical Bayes

In our model, the mixture weights, $\boldsymbol{\pi}$, are unknown. Following MASH [12], we use an empirical Bayes approach that combines both frequentist and Bayesian techniques. The first step, the frequentist step of empirical Bayes, involves maximizing the marginal likelihood to estimate $\hat{\boldsymbol{\pi}}$, which we do using stochastic approximation expectation maximization (SAEM) [13, 14]. The second step, the Bayesian step of empirical Bayes, uses $\hat{\boldsymbol{\pi}}$ as a plug-in estimate of $\boldsymbol{\pi}$ for downstream posterior sampling.

The ideal approach to obtain the MLE of $\boldsymbol{\pi}$ is to directly maximize the marginal likelihood

$$p\left(\mathbf{U}^\top \hat{\boldsymbol{\gamma}} \mid \boldsymbol{\pi}\right) = \int_{\mathbb{R}^{2M}} p\left(\mathbf{U}^\top \hat{\boldsymbol{\gamma}}, \boldsymbol{\gamma} \mid \boldsymbol{\pi}\right) d\boldsymbol{\gamma} .$$

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

However, this integral is intractable. Instead, we can use expectation maximization (EM) to maximize this marginal likelihood [15]. Suppose $\boldsymbol{\pi}^{(t)}$ represents the estimate at the tth iteration of the EM algorithm. The expectation step involves evaluating

$$Q\left(\boldsymbol{\pi} \mid \boldsymbol{\pi}^{(t)}\right) \triangleq \mathbb{E}_{\boldsymbol{\gamma} \sim p\left(\cdot \mid \mathbf{U}^\top \hat{\boldsymbol{\gamma}}, \boldsymbol{\pi} = \boldsymbol{\pi}^{(t)}\right)} \left[\log p\left(\mathbf{U}^\top \hat{\boldsymbol{\gamma}}, \boldsymbol{\gamma} \mid \boldsymbol{\pi}\right)\right] .$$

Since the analytical form of $Q\left(\boldsymbol{\pi} \mid \boldsymbol{\pi}^{(t)}\right)$ is also intractable under our model, we use SAEM to approximate the expectation. We sample $\boldsymbol{\gamma}^{(1)}, \ldots, \boldsymbol{\gamma}^{(B)} \overset{\text{iid}}{\sim} p\left(\cdot \mid \mathbf{U}^\top \hat{\boldsymbol{\gamma}}, \boldsymbol{\pi} = \boldsymbol{\pi}^{(t)}\right)$ using Hamiltonian Monte Carlo [16, 17]. Then, we approximate the expectation with

$$\mathbb{E}_{\boldsymbol{\gamma} \sim p\left(\cdot \mid \mathbf{U}^\top \hat{\boldsymbol{\gamma}}, \boldsymbol{\pi} = \boldsymbol{\pi}^{(t)}\right)} \left[\log p\left(\mathbf{U}^\top \hat{\boldsymbol{\gamma}}, \boldsymbol{\gamma} \mid \boldsymbol{\pi}\right)\right] \approx \frac{1}{B} \sum_{b=1}^{B} \log p\left(\mathbf{U}^\top \hat{\boldsymbol{\gamma}}, \boldsymbol{\gamma}^{(b)} \mid \boldsymbol{\pi} = \boldsymbol{\pi}^{(t)}\right) .$$

To speed up computation, we use stochastic gradient ascent, with each chromosome representing one mini-batch [18]. In the mini-batch approach, each expectation is a convex sum of the previous expectation and the current expectation. Let $(c_t)_{t \geq 1}$ be a positive, decreasing sequence such that $\sum_{t=1}^{\infty} c_t = \infty$ and $\sum_{t=1}^{\infty} c_t^2 < \infty$. We use

$$\widetilde{Q}\left(\boldsymbol{\pi} \mid \boldsymbol{\pi}^{(t)}\right) \triangleq (1 - c_t)\, \widetilde{Q}\left(\boldsymbol{\pi} \mid \boldsymbol{\pi}^{(t-1)}\right) + \frac{c_t}{B} \sum_{b=1}^{B} \log p\left(\mathbf{U}^\top \hat{\boldsymbol{\gamma}}, \boldsymbol{\gamma}^{(b)} \mid \boldsymbol{\pi} = \boldsymbol{\pi}^{(t)}\right)$$

in the expectation step of EM. To define a valid recursion, we set $\widetilde{Q}\left(\boldsymbol{\pi} \mid \boldsymbol{\pi}^{(0)}\right) = 0$. Maximization, the second step in EM, is performed using a fixed number of iterations of stochastic gradient ascent with a fixed step size to obtain

$$\boldsymbol{\pi}^{(t+1)} = \arg\max_{\boldsymbol{\pi}} \widetilde{Q}\left(\boldsymbol{\pi} \mid \boldsymbol{\pi}^{(t)}\right) .$$

Gradients were estimated using automatic differentiation [7, 8]. After some number of iterations, T, until convergence, we set $\hat{\boldsymbol{\pi}} = \boldsymbol{\pi}^{(T)}$.

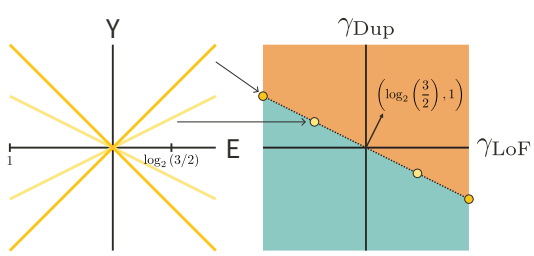

FIGURE E.1. The boundary between buffering curves is represented by the yellow gene dosage response curves on the left. These map to a diagonal across the burden effect plot. To estimate signal in one direction of the diagonal versus the other, we use the normal vector in black shown on the right.

We use Hamiltonian Monte Carlo [16, 17] to sample from the posterior distribution of $\boldsymbol{\gamma}$. Following the empirical Bayes approach, $\hat{\boldsymbol{\pi}}$ is used as a plug-in estimate. That is, we draw samples $\boldsymbol{\gamma}^{(1)}, \ldots, \boldsymbol{\gamma}^{(B)} \overset{\text{iid}}{\sim}$

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

$p\left(\cdot \mid \mathbf{U}^\top \hat{\boldsymbol{\gamma}}, \boldsymbol{\pi} = \hat{\boldsymbol{\pi}}\right)$. The estimand of interest is

$$\xi \triangleq \mathbb{E}\left[\log_2\left(\frac{3}{2}\right)\gamma_{1j} + \gamma_{2j} \,\middle|\, \mathbf{U}^\top \hat{\boldsymbol{\gamma}}\right] \approx \frac{1}{BM}\sum_{b=1}^{B}\sum_{j=1}^{M}\log_2\left(\frac{3}{2}\right)\gamma_{1j}^{(b)} + \gamma_{2j}^{(b)},$$

which is estimated using the posterior samples. The estimand represents the dot product of $\boldsymbol{\gamma}_{\cdot j}$ with the normal vector of the diagonal separating the two types of buffering within our model (Figure E.1). To estimate the buffering from monotone curves (ξ_{m}) versus non-monotone curves (ξ_{nm}), we use posterior samples that are only monotone or non-monotone respectively. That is,

$$\begin{aligned}
\xi_{\mathrm{nm}}\mathbb{P}\left(\gamma_{1j}\gamma_{2j} > 0\right) &= \mathbb{E}\left[\log_2\left(\frac{3}{2}\right)\gamma_{1j} + \gamma_{2j} \,\middle|\, \mathbf{U}^\top \hat{\boldsymbol{\gamma}}, \gamma_{1j}\gamma_{2j} > 0\right]\mathbb{P}\left(\gamma_{1j}\gamma_{2j} > 0\right) \\
&\approx \frac{1}{BM}\sum_{b=1}^{B}\sum_{\gamma_{1j}^{(b)}\gamma_{2j}^{(b)} > 0}\log_2\left(\frac{3}{2}\right)\gamma_{1j}^{(b)} + \gamma_{2j}^{(b)} \\
\xi_{\mathrm{m}}\mathbb{P}\left(\gamma_{1j}\gamma_{2j} > 0\right) &= \mathbb{E}\left[\log_2\left(\frac{3}{2}\right)\gamma_{1j} + \gamma_{2j} \,\middle|\, \mathbf{U}^\top \hat{\boldsymbol{\gamma}}, \gamma_{1j}\gamma_{2j} < 0\right]\mathbb{P}\left(\gamma_{1j}\gamma_{2j} < 0\right) \\
&\approx \frac{1}{BM}\sum_{b=1}^{B}\sum_{\gamma_{1j}^{(b)}\gamma_{2j}^{(b)} < 0}\log_2\left(\frac{3}{2}\right)\gamma_{1j}^{(b)} + \gamma_{2j}^{(b)}.
\end{aligned}$$

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

F Data Analysis

Synonymous Variants

Stratification or other confounding can inflate the estimated effect sizes from association analyses. We can ensure that we are adequately controlling for confounding by estimating the effect of synonymous variants, which are expected to not affect protein function and downstream traits.

We used synonymous variants to test the effect of using different subsets of individuals. The original analysis of LoF variants in the UKB [19] used a subset of around 430K individuals with genetic similarity to the EUR superpopulation from the 1000 Genomes Project. In addition to a population with genetic similarity to EUR (called EUR for brevity), we tested the subset of 390K unrelated individuals in the EUR subset (called unrelated) and the subset of all 460K individuals with whole-exome sequencing (WES) data (called WES). This allowed us to determine if relatedness or population stratification inflated our estimates of effect size. The EUR population was defined using self-reported information and boundaries in genotyping PC space from prior genetic analysis in the UKB [20]: $-20 \leq \mathrm{PC1} \leq 40$ and $-25 \leq \mathrm{PC2} \leq 10$ (Array items 1 and 2 from field 22009 in the UKB) for either self-identified White British or self-identified non-British White (Field 21000 in the UKB).

Additionally, we also used synonymous variants to test the effect of using different numbers of genotyping PCs. The original analysis used 10 genotyping PCs when performing association analyses within populations with high genetic similarity [19]. In addition to 10 genotyping PCs, we tested 15 and 20 genotyping PCs since we planned to use as many individuals in the UKB as possible, which might introduce additional confounding due to population stratification.

To test for stratification, we ran synonymous variant burden tests on a subset of nine continuous traits: height, body mass index (BMI), low-density lipoprotein cholesterol (LDL-C), mean corpuscular hemoglobin (MCH), red blood cell distribution width (RDW), forced vital capacity (FVC), creatinine, cystatin C, and the north coordinate of the place of birth in the United Kingdom (NC). Strong effects for NC should be a good measure of uncontrolled stratification.

Number of Genotyping Principal Components

Synonymous variants are expected to not have any effect on traits. Therefore, the mean squared effect of synonymous variant burden associations should provide an estimate of inflation in effect sizes due to other sources of confounding. The gold standard for genetic association analysis is using the cohort of unrelated individuals with high genetic similarity. Compared to this cohort, the amount of inflation was indistinguishable in the EUR and WES cohorts (Figure F.1). Thus, we decided to use the WES cohort to maximize our sample size.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

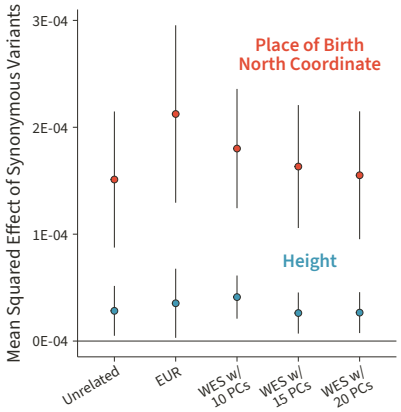

FIGURE F.1. The mean squared effect of synonymous variants for various cohorts of individuals and different numbers of genotyping PCs. The unrelated and EUR cohort burden tests used 15 genotyping PCs. We tested 10, 15, and 20 genotyping PCs in the WES cohort burden tests.

The effect of including 10 genotyping PCs was also indistinguishable from 15 or 20 genotyping PCs (Figure F.1). The original analysis included 10 genotyping PCs [19], but performed analyses in cohorts of high genetic similarity. Since we were including all individuals in the UKB, we decided to conservatively use 15 genotyping PCs.

INFLATION FROM CONFOUNDING

Since we detected a significant mean squared effect for NC, we were concerned about inflation of effect sizes due to confounding. To test the effect of this inflation, we compared the mean squared effect of LoF variant burden associations with the mean squared effect of synonymous variant burden associations across various traits. We noted that the mean squared effect of LoF burden associations was an order of magnitude larger than the effect of synonymous burden associations (Figure F.2). Thus, we concluded that although confounding is likely present in our association tests, the signal is at least 10 times greater than the bias.

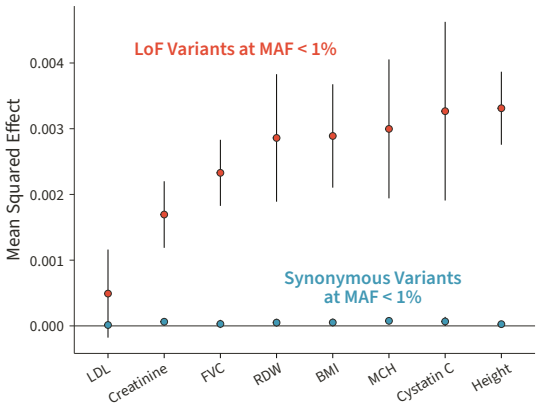

FIGURE F.2. LoF variants have a magnitude larger effect on traits than synonymous variants across various continuous traits.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

Utility of Misannotation Probability

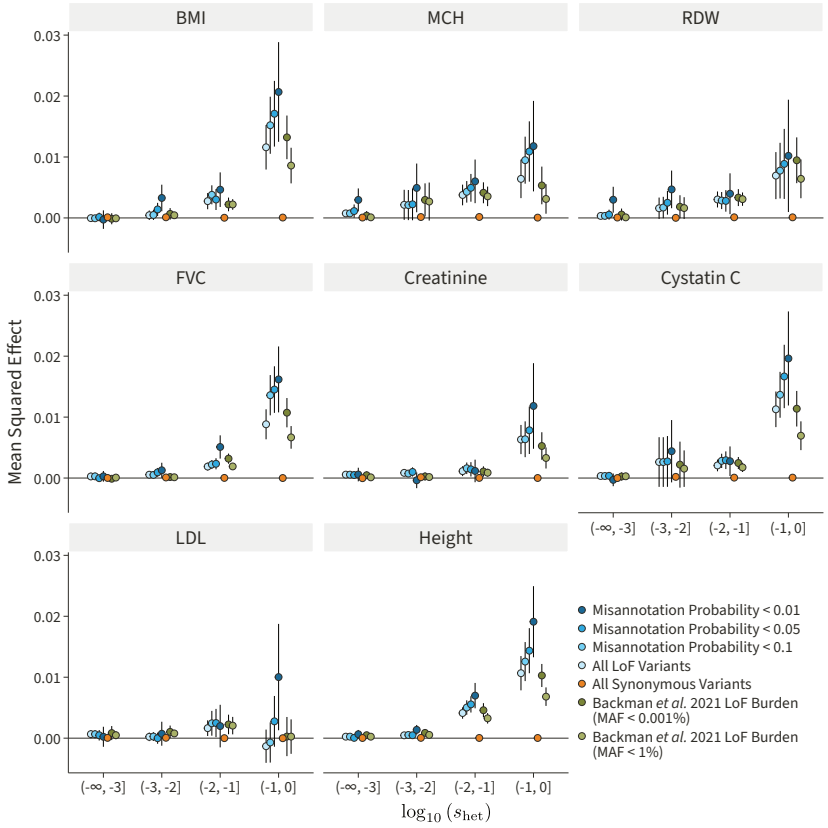

Figure F.3. The mean squared effect of genes in various buckets of $\log_{10}(s_{\text{het}})$ for various traits. Genes with larger effects tend to be more constrained. All data here uses a MAF < 1% filter other than the MAF < 0.001% associations from [19]. The use of the misannotation probability increases the signal detected. The signal in the original analysis of the LoF data is also shown for reference. 95% confidence intervals are displayed for each estimate.

Burden tests often use various maximum minor allele frequency (MAF) filters. For instance, Backman et al. used MAF filters of 1%, 0.1%, 0.01%, and 0.001% [19]. Presumably, such filters assume that increasingly stringent filters will increase the true positive LoF variants that are aggregated into the burden genotype, as LoF variants that are at high frequency in the population might represent misannotated LoF variants. However, such filters result in asymmetric behavior across genes depending on their constraint. For example, the highly stringent 0.001% filter will reduce the false-positive rate in highly constrained genes, but will remove true LoF variants in unconstrained genes. Ideally, a gene under high constraint should use a stringent MAF filter, while a gene under low constraint should use a liberal MAF filter.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

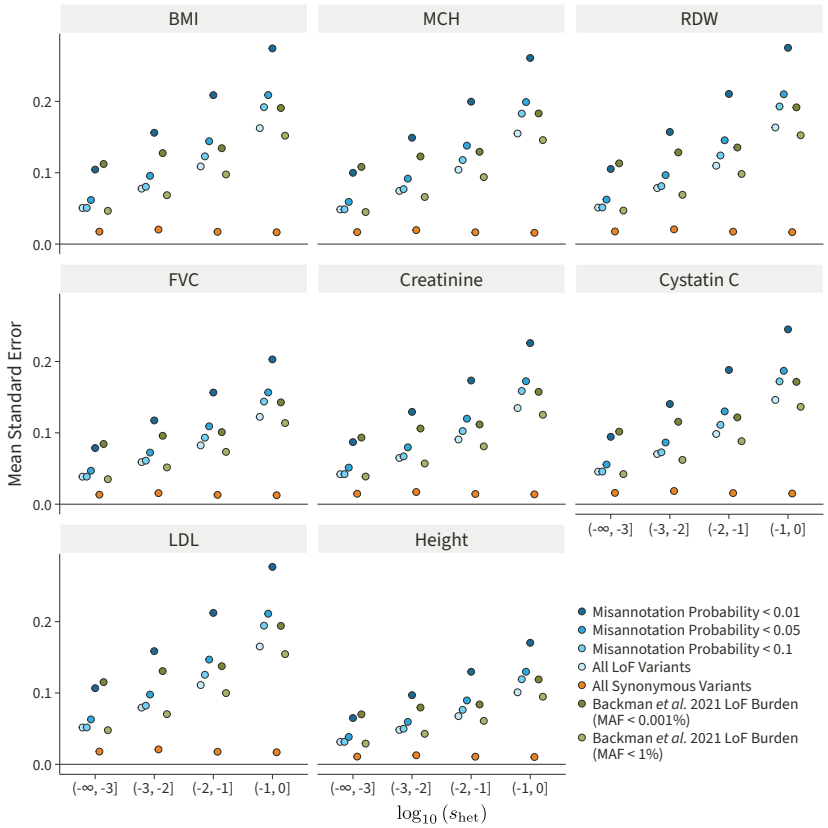

FIGURE F.4. The mean standard error of burden estimates for genes in various buckets of $\log_{10}(s_{\text{het}})$ for various traits. Since larger-effect variants are under increased selective constraint, their frequencies are lower and the estimation noise is larger. All data here uses a MAF < 1% filter other than the MAF < 0.001% associations from [19]. The use of the misannotation probability increases the noise of the estimates. The noise in the original analysis of the LoF data is also shown for reference. 95% confidence intervals are present on the plots but not visible due to their small length.

To account for this, Zeng et al. have calculated misannotation probabilities for all potential single nucleotide polymorphisms (SNPs) in genes for which they had estimated s_{het} values [21]. We used various misannotation probability filters instead of MAF filters. We tested filters of 10%, 5%, and 1% misannotation probability for all variants with MAF < 1%. We estimated the total signal, as measured by the mean squared effect size, in various buckets of gene constraint (Figure F.3). Increasingly stringent misannotation probability filters increased the signal across various selection buckets. In addition, the various filters were as or more effective than both the 1% and 0.001% MAF filters from [19].

Increasing stringency with the misannotation probability filters does increase the amount of estimation noise in $\hat{\gamma}_{\text{LoF}}$ as more LoF variants are removed from the burden genotypes (Figure F.4). We found that a misannotation probability filter of 10% provided signal comparable to or better than a MAF filter of 0.001% with a minimal increase in noise. This filter was used in all subsequent analyses.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

CNV Genotyping Error

When we first estimated the LD matrices for the duplication burden genotypes, we noticed large amounts of long-range LD (Figure F.5).

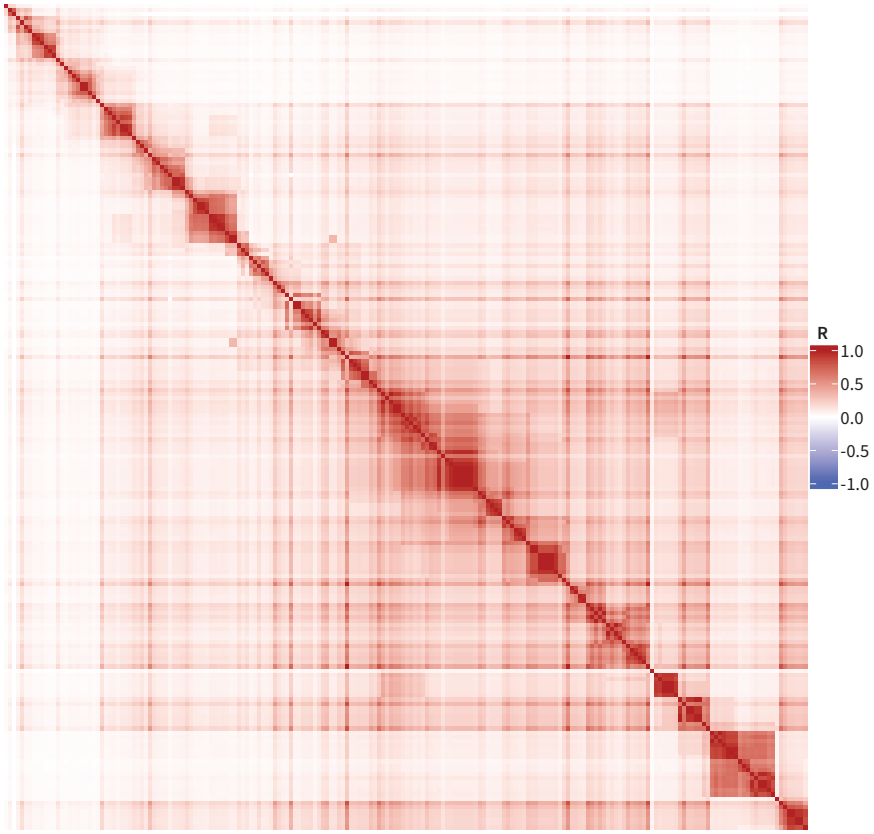

Figure F.5. The correlation matrix for the first 200 genes on chromosome 3 using all the samples. Although segregating duplications can be large, they are not expected to capture so many genes.

We reasoned that this is driven by a few poor-quality samples with systematically higher copy number. In their initial report on the UKB whole-genome sequencing (WGS) data, Li et al. reported an average structural variant (SV) burden of 3.6 Mbp per haploid genome [22]. Assuming that the total copy number variant (CNV) burden is on a similar order of magnitude, we decided to exclude any samples with more than 10 Mbp of either deletion or duplication burden. Most samples had less than 10 Mbp of duplications or deletions. However, some samples had upwards of 100 Mbp of affected sequence (Figure F.6).

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

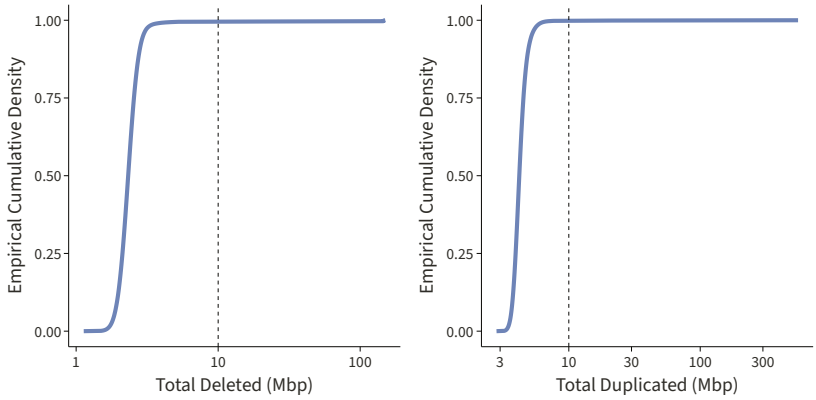

FIGURE F.6. The empirical cumulative density function (eCDF) of the total amount of genome sequence affected by either deletion or duplication. Our 10 Mbp filtering step impacted only a few samples.

Filtering out 2,674 samples with deletion or duplication burden greater than 10 Mbp resulted in more reasonable duplication LD estimates (Figure F.7).

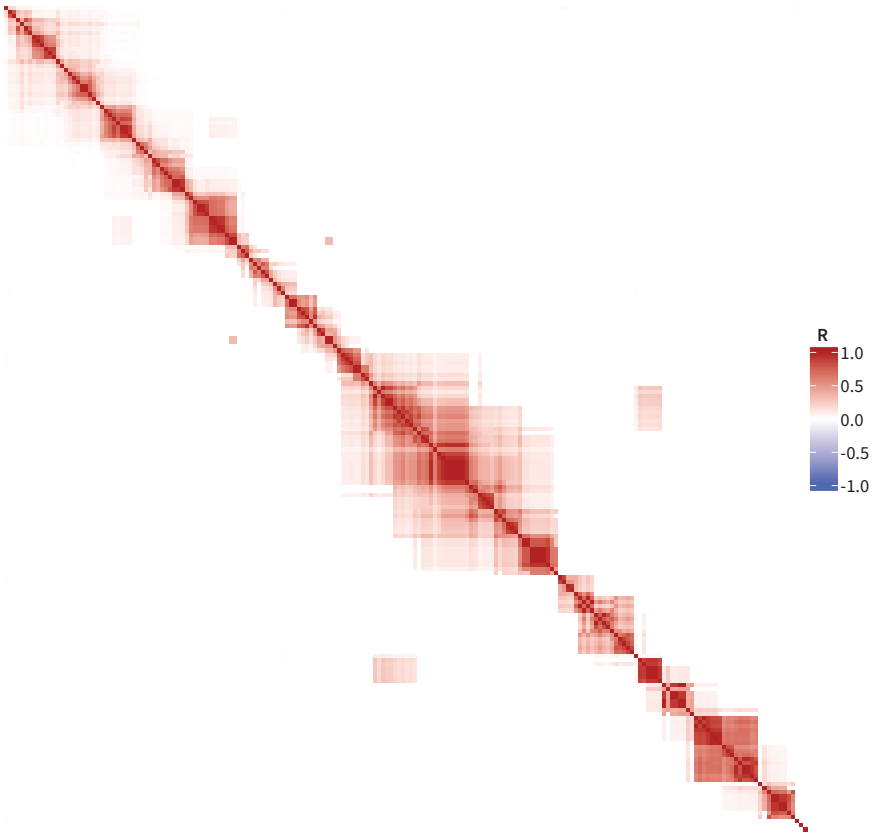

FIGURE F.7. The correlation matrix for the first 200 genes on chromosome 3 using the filtered set of samples.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

Estimating CNV Linkage Disequilibrium

Multiple covariates need to be accounted for when performing burden tests. Suppose that $\mathbf{Z}_c \in \mathbb{R}^{N \times C}$ represents the covariate values for the deletion or duplication burden tests for C covariates. The burden test model used to derive summary statistics in REGENIE is

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\gamma}_2 + \mathbf{Z}_c\boldsymbol{\lambda}_c + \boldsymbol{\epsilon}_2 \,,$$

where $\boldsymbol{\lambda}_c \in \mathbb{R}^C$ are the effects of the covariates on the phenotype. To remove these effects before estimating the LD matrix, we partial out the covariates by projecting the burden genotypes into the subspace orthogonal to the column space of the covariates. Specifically, we define

$$\mathbf{Z}' \triangleq \mathbf{Z} - \mathbf{Z}_c \left(\mathbf{Z}_c^\top \mathbf{Z}_c\right)^{-1} \mathbf{Z}_c^\top \mathbf{Z} \,,$$

which removes the effect of the covariates by removing their contribution to the burden genotypes using the projection matrix $\mathbf{Z}_c \left(\mathbf{Z}_c^\top \mathbf{Z}_c\right)^{-1} \mathbf{Z}_c^\top$. Then, we estimate the correlation matrix using these augmented burden genotypes $\mathbf{Z}'$.

Significant Average Effects

We estimated average burden effects using unbiased estimators. Traits with significant LoF and duplication burden effects were ascertained at an $\alpha = 0.05$ level (Table F.1).

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

Trait	$\hat{\bar{\gamma}}_1 \left(\times 10^2\right)$	$\hat{\bar{\gamma}}_1$ 95% CI	$\hat{\bar{\gamma}}_2 \left(\times 10^2\right)$	$\hat{\bar{\gamma}}_2$ 95% CI
Alkaline phosphatase	1.04	[0.79, 1.29]	0.59	[0.25, 0.93]
Cystatin C	1.02	[0.78, 1.25]	1.27	[0.94, 1.59]
White blood cell count	0.92	[0.65, 1.19]	0.53	[0.16, 0.89]
Neutrophil count	0.84	[0.57, 1.11]	0.52	[0.15, 0.89]
C-reactive protein	0.84	[0.57, 1.11]	0.53	[0.16, 0.90]
Body mass index (BMI)	0.79	[0.53, 1.05]	0.77	[0.41, 1.13]
Waist circumference	0.78	[0.54, 1.01]	0.81	[0.49, 1.14]
Pulse rate	0.70	[0.42, 0.98]	0.61	[0.23, 0.99]
Body fat percentage	0.64	[0.44, 0.84]	0.49	[0.21, 0.76]
Glycated haemoglobin (HbA1c)	0.55	[0.30, 0.81]	0.48	[0.13, 0.83]
Hip circumference	0.45	[0.19, 0.72]	0.50	[0.14, 0.87]
Vitamin D	-0.46	[-0.74, -0.18]	-0.65	[-1.03, -0.27]
Apolipoprotein A	-0.64	[-0.90, -0.38]	-0.65	[-1.00, -0.31]
HDL cholesterol	-0.68	[-0.93, -0.43]	-0.69	[-1.03, -0.36]
Seated height	-0.74	[-0.94, -0.53]	-0.34	[-0.62, -0.07]
Standing height	-0.81	[-0.98, -0.65]	-0.48	[-0.70, -0.26]
Hand grip strength (right)	-0.85	[-1.04, -0.66]	-0.78	[-1.04, -0.52]
Hand grip strength (left)	-0.85	[-1.04, -0.66]	-0.65	[-0.91, -0.39]
FEV1	-0.96	[-1.17, -0.75]	-0.96	[-1.24, -0.68]
Forced vital capacity (FVC)	-0.98	[-1.18, -0.78]	-0.84	[-1.11, -0.56]
Peak expiratory flow (PEF)	-1.04	[-1.27, -0.80]	-0.99	[-1.31, -0.67]

TABLE F.1. Traits with significant average burden effects were ascertained at an $\alpha = 0.05$ level. All such traits have non-monotone average effects. These estimates are from unbiased estimators with exact confidence intervals.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

Lack of Average Effect in Genome-Wide Association Studies

For traits with significant average burden effects (Table F.1), we downloaded genome-wide association study (GWAS) summary statistics from the Neale Lab (https://www.nealelab.is/uk-biobank, version 3, both sexes) based on 337,199 White British individuals in the UKB. In their processing pipeline, phenotypic values were inverse-rank normal transformed. Age, age squared, inferred sex, age-by-sex, age-squared-by-sex, and 20 genotyping PCs were included as covariates.

We estimated conditionally independent hits and their effect sizes using GCTA-COJO [23] with parameters `--cojo-p 5e-8 --cojo-slct`. We used the genotypes of 10,000 unrelated White British individuals in the UKB to compute the LD reference panel. We polarized significant trait-associated variants into ancestral and derived states with variation features obtained from the Ensembl Variation database [24].

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

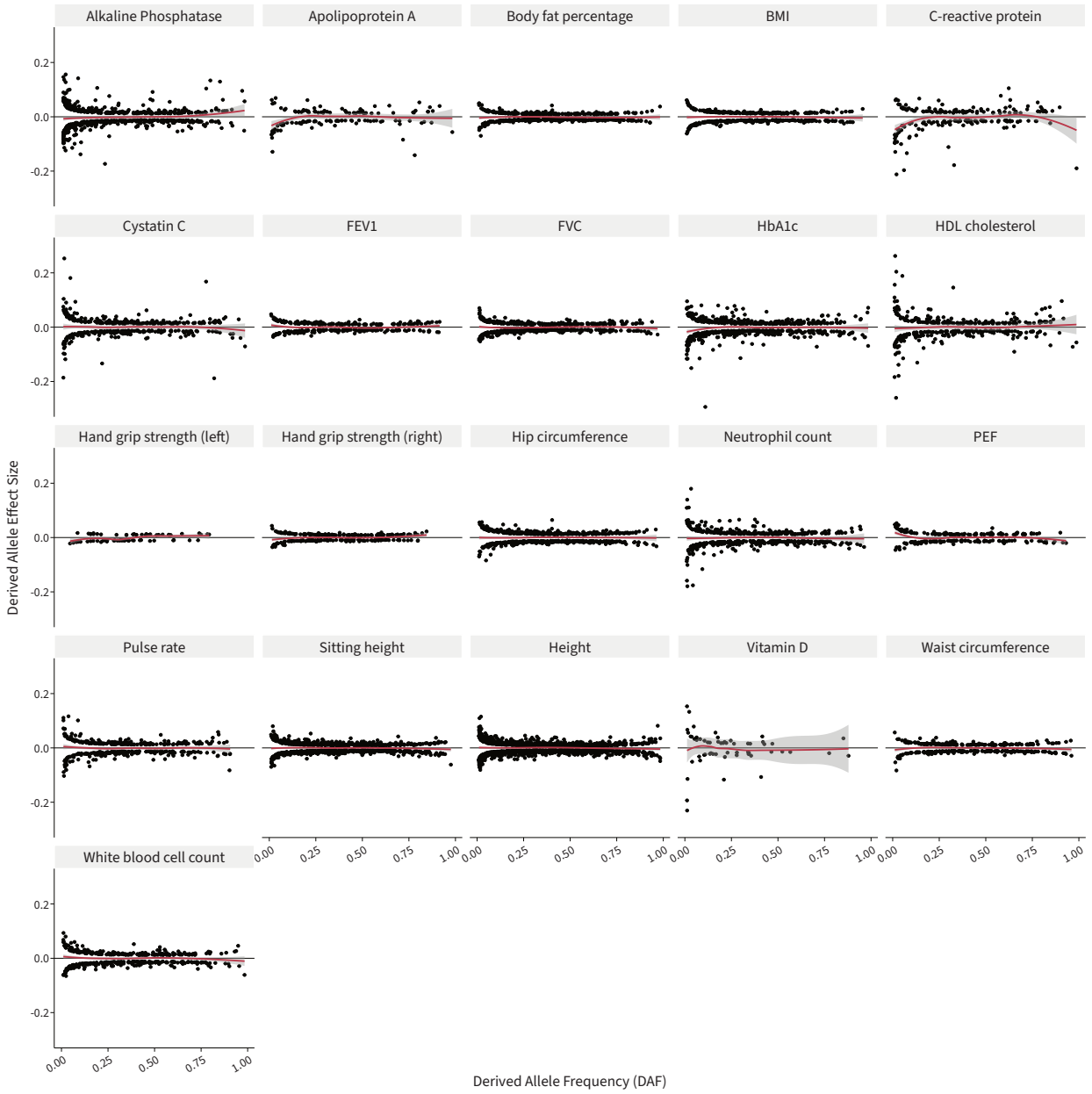

FIGURE F.8. Trumpet plots of the derived allele effect size versus the derived allele frequency for top genome-wide association study (GWAS) associations for traits with significant loss-of-function (LoF) and duplication average burden effects.

In the trumpet plots of derived allele effect versus derived allele frequency (Figures 1B and F.8), we use top significant conditionally independent hits per locus detected in a GWAS. One concern in such a plot is that some top hits may be tagging SNPs, and the causal SNPs may be in negative or positive LD with the tagging SNPs. Tagging SNPs in negative LD with the causal SNP can mask an underlying true non-zero average effect.

To address these concerns, we point to recent work done to understand selection using common variation [25]. Under neutral coalescent simulations, the causal and tagging SNPs are expected to be in

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

positive LD with high probability for derived allele frequencies in the interval $[0, 0.5]$ as shown in Figure S14 of [25]. The trumpet plot continues to display a nearly zero average effect in this interval, suggesting that the underlying causal SNPs also do not show an average effect on the trait.

Analysis of Monotonicity Effect Estimates

To better understand the distribution of monotonicity estimates, we used a hierarchical model to infer the distribution of effects. Let $\left\{\hat{\phi}_i\right\}_{i=1}^{T}$ represent the observed monotonicity estimates for T traits with standard errors $\hat{c}_i = \widehat{\mathrm{SE}}\left(\hat{\phi}_i\right)$. Suppose that

$$\hat{\phi}_i \mid \phi_i, \hat{c}_i \sim \mathcal{N}\left(\phi_i, \hat{c}_i^2\right) .$$

We assume that ϕ_i is drawn from a transformed beta distribution over the interval $[-1, 1]$,

$$\phi_i \overset{\text{iid}}{\sim} \mathrm{TransformedBeta}\left(\alpha, \beta, [-1, 1]\right) .$$

The likelihood for the transformed beta distribution can be derived using the transformation of random variables approach. Suppose that $U \sim \mathrm{Beta}\left(\alpha, \beta\right)$. We define V as

$$V = g\left(U\right) = (b - a)U + a$$

for $a, b \in \mathbb{R}$ and $a < b$. Note that the support of V is $[a, b]$. We then define

$$V \sim \mathrm{TransformedBeta}\left(\alpha, \beta, [a, b]\right) .$$

The inverse transform is

$$g^{-1}\left(V\right) = \frac{V - a}{b - a}$$

and the derivative is

$$\frac{d}{dV} g^{-1}\left(V\right) = \frac{1}{b - a} .$$

The likelihood is therefore

$$\begin{aligned}\mathrm{TransformedBeta}\left(v\,; \alpha, \beta, [a, b]\right) &= \mathrm{Beta}\left(g^{-1}\left(v\right)\,; \alpha, \beta\right) \left|\frac{d}{dv} g^{-1}\left(v\right)\right| \\ &= \frac{1}{b - a} \mathrm{Beta}\left(\frac{v - a}{b - a}\,; \alpha, \beta\right) .\end{aligned}$$

We use an empirical Bayes approach to fit the prior distribution. We do so by maximizing the evi-

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

dence,

$$
\begin{aligned}
\hat{\alpha}, \hat{\beta} &= \arg\max_{\alpha,\beta} \prod_{i=1}^{T} \mathbb{P}\left(\hat{\phi}_i, \alpha, \beta \mid \hat{c}_i\right) \\
&= \arg\max_{\alpha,\beta} \prod_{i=1}^{T} \int_{-1}^{1} \mathbb{P}\left(\hat{\phi}_i \mid \phi_i, \hat{c}_i\right) \mathbb{P}\left(\phi_i \mid \alpha, \beta\right) \, d\phi_i \\
&= \arg\max_{\alpha,\beta} \prod_{i=1}^{T} \int_{-1}^{1} \mathcal{N}\left(\hat{\phi}_i \,; \phi_i, \hat{c}_i^2\right) \mathrm{TransformedBeta}\left(\phi_i \,; \alpha, \beta, [-1, 1]\right) \, d\phi_i \,.
\end{aligned}
$$

We used Riemann integration to numerically evaluate the evidence and used gradient ascent with a fixed number of steps to estimate $\hat{\alpha}$ and $\hat{\beta}$. Gradients were estimated using automatic differentiation [7, 8].

We estimated $\hat{\phi}$ using both the MLE and the MoM estimates. The main results for the MLEs are presented in Figure 2B. The point estimates and confidence intervals for the MLEs are displayed in Figure F.9.

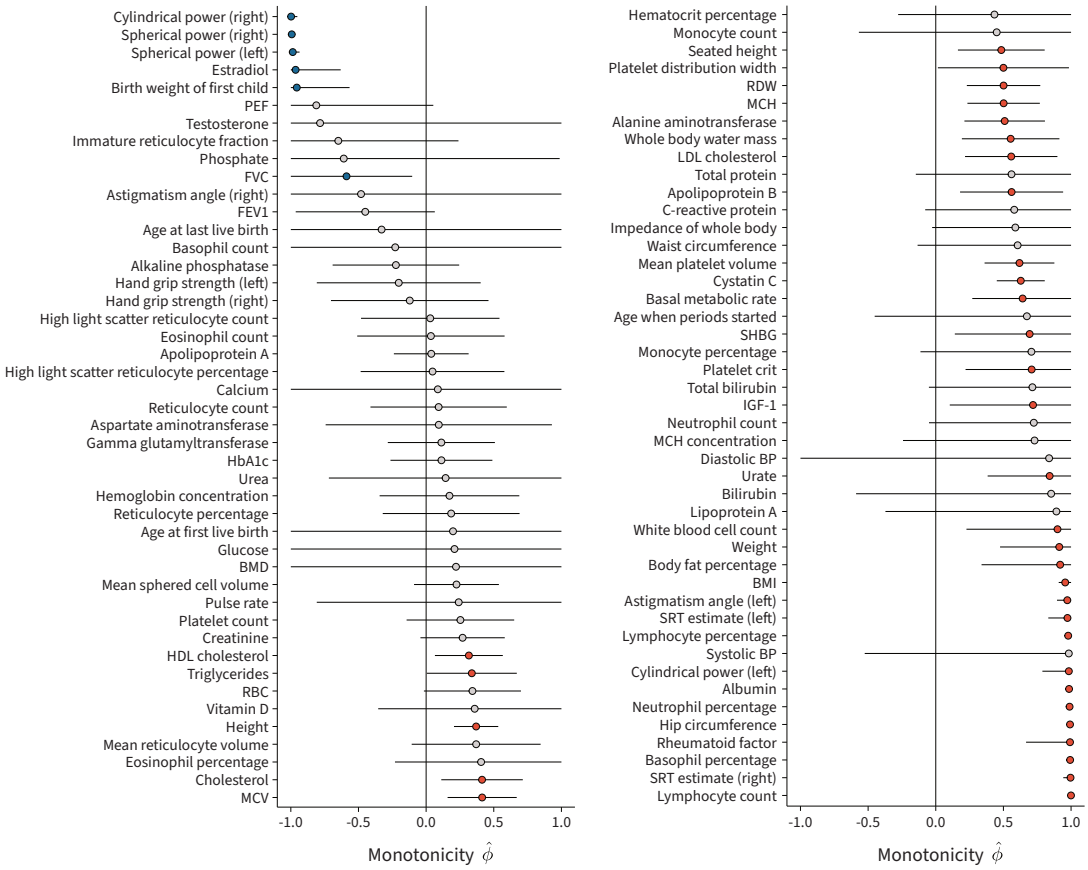

FIGURE F.9. Point estimates with 95% confidence intervals for monotonicity from the maximum likelihood estimation approach. Blue points represent significantly negative monotonicity, while red points represent significantly positive monotonicity. Grey points represent non-significant values.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

The same analysis was conducted with the MoM estimates, which provided a similar support for monotonicity across traits (Figure F.10). 99.96% of the density of the inferred effect size distribution was over the positive part of the domain of ϕ. For this analysis, we removed traits with estimates greater than one. We also excluded traits where the estimate was undefined, which can occur with negative estimates of $\hat{\sigma}^2_{11}$ or $\hat{\sigma}^2_{22}$. The point estimates and confidence intervals for the MoM estimates are displayed in Figure F.11.

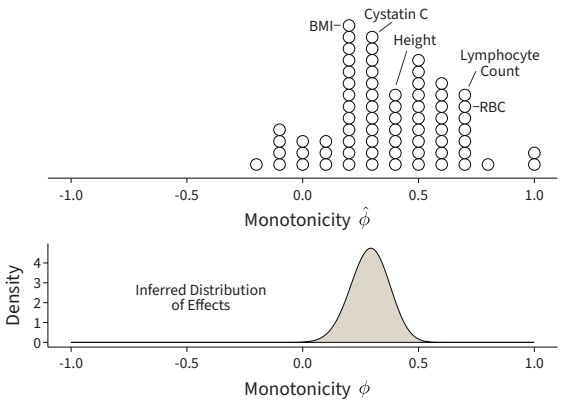

FIGURE F.10. (Top) The method-of-moments (MoM) point estimates for monotonicity for the traits that were analyzed. (Bottom) The inferred distribution of effects from a random-effects meta-analysis.

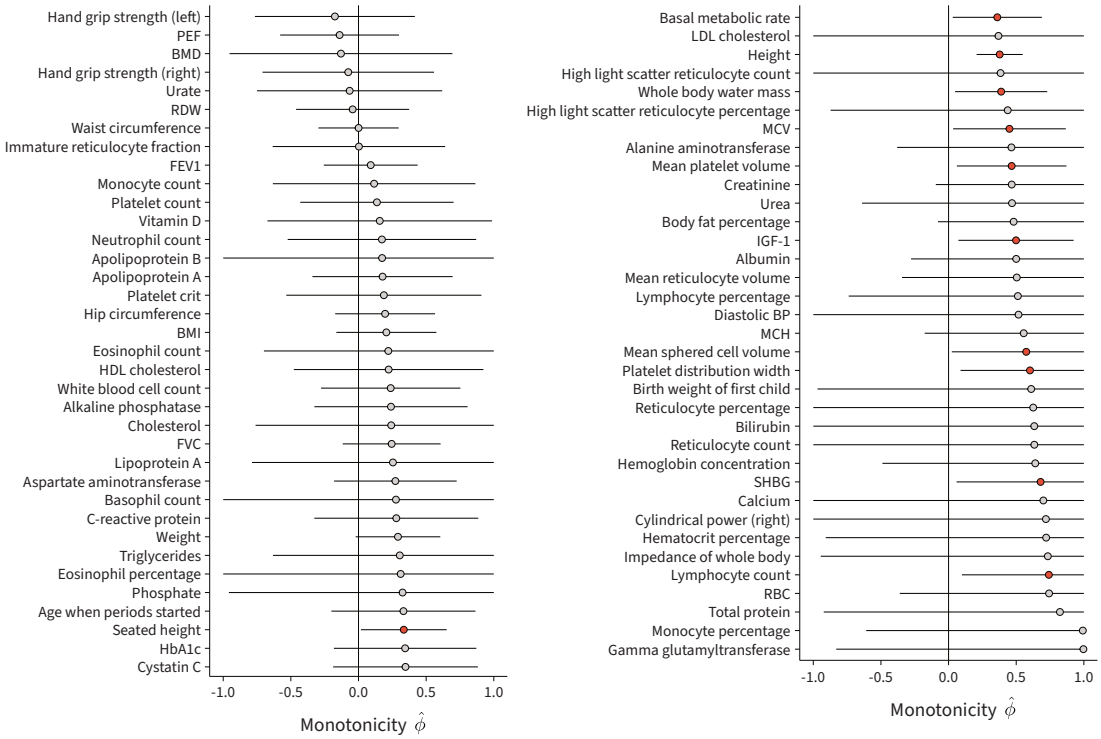

FIGURE F.11. Point estimates with 95% confidence intervals for monotonicity from the method-of-moments (MoM) approach. Red points represent significantly positive monotonicity. Grey points represent non-significant values.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

Concordance of LoF and Deletion Effects

To estimate the extent to which the burden tests using LoF variants are concordant with burden tests performed using deletions, we estimated the genetic correlation between the effects from the two variant classes. We did so by performing maximum likelihood estimation using the monotonicity model. When we assume that the average effects are zero ($\overline{\gamma}_1 = \overline{\gamma}_2 = 0$), the correlation

$$\rho = \frac{\sigma_{12}}{\sigma_{11}\sigma_{22}}$$

is similar to the genetic correlation derived under the LDSC model [26]. While we assume a normal prior distribution over the latent effect sizes, the original LDSC model assumes no functional prior form and only specifies finite first and second moments.

We derived the MLE $\hat{\rho}$ using the same inference machinery used for monotonicity, assuming fixed zero average effect sizes. Approximate confidence intervals were estimated using the delta method.

Next, similar to the monotonicity analysis, we estimated the latent distribution of genetic correlations by fitting a transformed beta distribution. That is, let $\{\hat{\rho}_i\}_{i=1}^T$ represent the observed genetic correlation estimates for T traits with standard errors $\hat{c}_i = \widehat{\mathrm{SE}}\left(\hat{\rho}_i\right)$. Suppose that

$$\hat{\rho}_i \mid \rho_i, \hat{c}_i \sim \mathcal{N}\left(\rho_i, \hat{c}_i^2\right) .$$

We assume that ρ_i is drawn from a transformed beta distribution over the interval $[-1, 1]$,

$$\rho_i \overset{\text{iid}}{\sim} \mathrm{TransformedBeta}\left(\alpha, \beta, [-1, 1]\right) .$$

We estimate $\hat{\alpha}$ and $\hat{\beta}$ using the same strategy as the monotonicity estimates.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

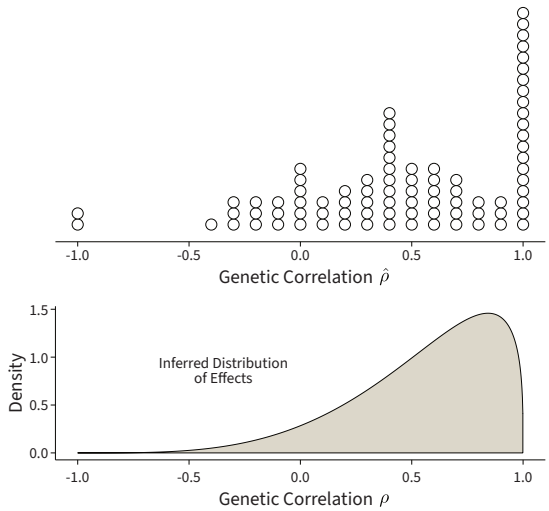

FIGURE F.12. (Top) The maximum likelihood estimates (MLEs) for genetic correlation for the traits that were analyzed. The genetic correlation is estimated between loss-of-function (LoF) burden effects and deletion burden effects. (Bottom) The inferred distribution of effects from a random-effects meta-analysis.

This analysis demonstrates that LoF variant and deletion burden tests are markedly similar (Figure F.12). 93.53% of the density of the inferred distribution of genetic correlations is over the positive part of the domain of ρ. For this analysis, we removed traits with $\hat{c}_i < 10^{-6}$ as these might represent unstable estimates from the optimization procedure. The estimated genetic correlation for these estimates was $\hat{\rho} = 1$, meaning that dropping them from our analysis biased our inferred distribution towards zero.

APPENDIX REFERENCES

[1] Chiara Auwerx et al. "The individual and global impact of copy-number variants on complex human traits". *The American Journal of Human Genetics* 109.4 (Apr. 2022), pp. 647–668. DOI: 10.1016/j.ajhg.2022.02.010.

[2] Benjamin H Good. "Linkage disequilibrium between rare mutations". *Genetics* 220.4 (Apr. 2022), iyac004. DOI: 10.1093/genetics/iyac004.

[3] Xiang Zhu and Matthew Stephens. "Bayesian large-scale multiple regression with summary statistics from genome-wide association studies". *The Annals of Applied Statistics* 11.3 (Sept. 2017), pp. 1561–1592. DOI: 10.1214/17-AOAS1046.

[4] Shun-ichi Amari. "Natural Gradient Works Efficiently in Learning". *Neural Computation* 10.2 (Feb. 1998), pp. 251–276. DOI: 10.1162/089976698300017746.

[5] Luigi Malagò and Giovanni Pistone. "Information Geometry of the Gaussian Distribution in View of Stochastic Optimization". *Proceedings of the 2015 ACM Conference on Foundations of Genetic Algorithms XIII*. FOGA '15. New York, NY, USA: Association for Computing Machinery, Jan. 2015, pp. 150–162. DOI: 10.1145/2725494.2725510.

[6] Larry Armijo. "Minimization of functions having Lipschitz continuous first partial derivatives." *Pacific Journal of Mathematics* 16.1 (Jan. 1966), pp. 1–3.

[7] Roy Frostig, Matthew James Johnson, and Chris Leary. "Compiling machine learning programs via high-level tracing". *SysML*. Stanford, CA, Feb. 2018.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.11.24317065; this version posted November 11, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

[8] Charles C. Margossian. "A Review of automatic differentiation and its efficient implementation". *WIREs Data Mining and Knowledge Discovery* 9.4 (Nov. 2018), e1305. DOI: 10.1002/WIDM.1305.

[9] Brendan K. Bulik-Sullivan et al. "LD Score regression distinguishes confounding from polygenicity in genome-wide association studies". *Nature Genetics* 47.3 (Feb. 2015), pp. 291–295. DOI: 10.1038/ng.3211.

[10] Christopher C Chang et al. "Second-generation PLINK: rising to the challenge of larger and richer datasets". *GigaScience* 4.1 (Dec. 2015), s13742–015–0047–8. DOI: 10.1186/s13742-015-0047-8.

[11] Daniel J. Weiner et al. "Polygenic architecture of rare coding variation across 394,783 exomes". *Nature* 614.7948 (Feb. 2023), pp. 492–499. DOI: 10.1038/s41586-022-05684-z.

[12] Sarah M. Urbut et al. "Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions". *Nature Genetics* 51.1 (Nov. 2018), pp. 187–195. DOI: 10.1038/s41588-018-0268-8.

[13] Bernard Delyon, Marc Lavielle, and Eric Moulines. "Convergence of a Stochastic Approximation Version of the EM Algorithm". *The Annals of Statistics* 27.1 (Feb. 1999), pp. 94–128.

[14] Estelle Kuhn and Marc Lavielle. "Coupling a stochastic approximation version of EM with an MCMC procedure". *ESAIM: Probability and Statistics* 8 (Aug. 2004), pp. 115–131. DOI: 10.1051/ps:2004007.

[15] A. P. Dempster, N. M. Laird, and D. B. Rubin. "Maximum Likelihood from Incomplete Data Via the EM Algorithm". *Journal of the Royal Statistical Society: Series B (Methodological)* 39.1 (Sept. 1977), pp. 1–22. DOI: 10.1111/j.2517-6161.1977.tb01600.x.

[16] Eli Bingham et al. "Pyro: Deep Universal Probabilistic Programming". *Journal of Machine Learning Research* 20.28 (Feb. 2019), pp. 1–6.

[17] Du Phan, Neeraj Pradhan, and Martin Jankowiak. *Composable Effects for Flexible and Accelerated Probabilistic Programming in NumPyro*. Dec. 2019. DOI: 10.48550/arXiv.1912.11554.

[18] Estelle Kuhn, Catherine Matias, and Tabea Rebafka. "Properties of the stochastic approximation EM algorithm with mini-batch sampling". *Statistics and Computing* 30.6 (Nov. 2020), pp. 1725–1739. DOI: 10.1007/s11222-020-09968-0.

[19] Joshua D. Backman et al. "Exome sequencing and analysis of 454,787 UK Biobank participants". *Nature* 599.7886 (Oct. 2021), pp. 628–634. DOI: 10.1038/s41586-021-04103-z.

[20] Nasa Sinnott-Armstrong et al. "Genetics of 35 blood and urine biomarkers in the UK Biobank". *Nature Genetics* 53.2 (Jan. 2021), pp. 185–194. DOI: 10.1038/s41588-020-00757-z.

[21] Tony Zeng et al. "Bayesian estimation of gene constraint from an evolutionary model with gene features". *Nature Genetics* (July 2024), pp. 1–12. DOI: 10.1038/s41588-024-01820-9.

[22] Shuwei Li et al. *Whole-genome sequencing of half-a-million UK Biobank participants*. Dec. 2023. DOI: 10.1101/2023.12.06.23299426.

[23] Jian Yang et al. "Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits". *Nature Genetics* 44.4 (Apr. 2012), pp. 369–375. DOI: 10.1038/ng.2213.

[24] Fergal J Martin et al. "Ensembl 2023". *Nucleic Acids Research* 51.D1 (Jan. 2023), pp. D933–D941. DOI: 10.1093/nar/gkac958.

[25] Roshni A. Patel et al. *Conditional frequency spectra as a tool for studying selection on complex traits in biobanks*. June 2024. DOI: 10.1101/2024.06.15.599126.

[26] Brendan Bulik-Sullivan et al. "An atlas of genetic correlations across human diseases and traits". *Nature Genetics* 47.11 (Sept. 2015), pp. 1236–1241. DOI: 10.1038/ng.3406.