

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

1

2 The genome-wide burdens of deletions, loss-of-function mutations, and duplications correlate with
3 many traits. Curiously, for most of these traits, variants that decrease expression have the same genome-
4 wide average direction of effect as variants that increase expression. This seemingly contradicts the
5 intuition that, at individual genes, reducing expression should have the opposite effect on a phenotype
6 as increasing expression. To understand this paradox, we introduce a concept called the gene dosage
7 response curve (GDRC) that relates changes in gene expression to expected changes in phenotype. We
8 show that, for many traits, GDRCs are systematically biased in one trait direction relative to the other
9 and, surprisingly, that as many as 40% of GDRCs are non-monotone, with large increases and decreases
10 in expression affecting the trait in the same direction. We develop a simple theoretical model that ex-
11 plains this bias in trait direction. Our results have broad implications for complex traits, drug discovery,
12 and statistical genetics.

13

14 ¹Department of Genetics, Stanford University, CA, USA

15 ²Department of Biology, Stanford University, CA, USA

16 *Corresponding authors: nmilind@stanford.edu, jspence@stanford.edu, pritch@stanford.edu

17 INTRODUCTION

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

21 One source of variation in gene expression is gain or loss of functional copies of a gene [9]. Such
22 gene dosage changes can have phenotypic consequences, with aneuploidies being an extreme example
23 [10–13]. Copy number variants (CNVs), which typically perturb the dosage of a few genes, are another
24 source of expression variation [14]. CNVs segregate in humans [15–18] and have measurable effects on
25 various traits and disorders [19–27], presumably by varying the expression of one or more of the genes
26 with atypical copy numbers [9].

27 There has been substantial work associating the total genome-wide CNV burden to traits [19–22, 24,
2 25, 27–29]. An individual’s genome-wide burden is estimated by counting the total number of CNVs car-
29 ried by an individual, regardless of which genes are affected. Associating this burden with a trait can be
30 thought of as roughly estimating how much deleting or duplicating a random gene affects the trait. The
31 large number of traits significantly associated with genome-wide deletion burden suggests that the ef-
32 fect of reducing the expression of a randomly chosen gene is often biased in a particular trait direction.
33 Interestingly, genome-wide duplication burden and genome-wide deletion burden often have the same
34 direction of effect [20–22]. That is, increasing expression has the same effect as decreasing expression on
35 average (Appendix A). This seemingly contradicts the intuition that, for a given gene, increasing expres-
36 sion should have the opposite effect of decreasing expression.

37 To make sense of these genome-wide CNV burden test results from the perspective of what is happen-
3 ing at individual genes, we estimated the gene-level effects of loss-of-function (LoF) variants, deletions,
39 and duplications using whole-exome sequencing (WES) and whole-genome sequencing (WGS) data from
40 the UK Biobank (UKB) [30, 31].

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

47 RESULTS

4 GENE-LEVEL BURDEN TESTS FOR LOSS-OF-FUNCTION VARIANTS AND DUPLICATIONS

49 To better understand how variants with different effects on dosage impact traits, we performed burden
50 tests for 410 continuous traits using LoF variants, deletions, and duplications in the UKB. In the analysis
51 presented, we focused on LoF variant burden tests because they are better powered than deletions, but
52 deletions showed broadly similar patterns of effect (Appendix F). We took care to keep the analysis of the
53 different variant types as similar as possible so that the burden effect sizes could be interpreted jointly.
54 We used the gene burden test implemented in REGENIE [32]. Continuous traits were selected from vari-
55 ous blood biochemistry, blood count, blood metabolite, and anthropometric measurements available in
56 the UKB (Data Availability). We accounted for sex, age, batch effects, and other sources of confounding by
57 including appropriate covariates (Methods) and performed quality control steps (Appendix F). The sum-
58 mary statistics from these burden tests are provided as a resource to the community (Data Availability).
59 We selected a subset of 91 continuous traits for analysis in the paper (Data Availability).

60 GENE DOSAGE RESPONSE CURVES COULD EXPLAIN PARADOXICAL PATTERNS IN BURDEN TESTS

61 Estimating the effect of genes on complex traits is central to understanding the mechanisms underlying
62 trait biology. Changes in gene expression, which we refer to here as ‘gene dosage’, are often assumed to
63 have a linear effect on a trait (Figure 1A). For example, transcriptome-wide association studies (TWASs)
64 use a linear model for the relationship between predicted expression and trait value [5, 6]. However,
65 LoF and duplication burden tests assess phenotypic effects at opposite ends of the dosage spectrum and
66 under a linear model would have opposite directions of effect. Thus, a linear model is incompatible with
67 the genome-wide burden effects of LoF variants and duplications having the same sign (Appendix A).
68 Recent work on the effects of gene dosage on molecular traits shows these relationships are often non-
69 linear [33, 34]; as we will show below, an assumption of nonlinearity is necessary but not sufficient to
70 explain these observations.

71 We next extended the observations of directionality by estimating the effect of each individual gene
72 on the trait using burden tests, and then averaging those estimates across all genes for a trait. We found
73 that the estimated *average burden effect* for LoF variants was almost always in the same direction as that
74 of duplications (Figure 1B). Curiously, however, when we look at top hits from GWASs, we find no evidence
75 for such an effect (Figures 1C and F.8).

76 Taken together, these observations provide contradictory information about the relationship be-
77 tween dosage and trait. To explain these observations, we need to understand the gene-level hetero-
78 geneity present in the dosage-response relationships. We envision the relationship between gene ex-
79 pression and trait as a gene dosage response curve (GDRC): the continuous relationship between gene
80 expression and average trait value. The core idea is that if a gene is involved in the biology of a trait, then
81 different baseline expression values for the gene will result in different trait values on average. A given

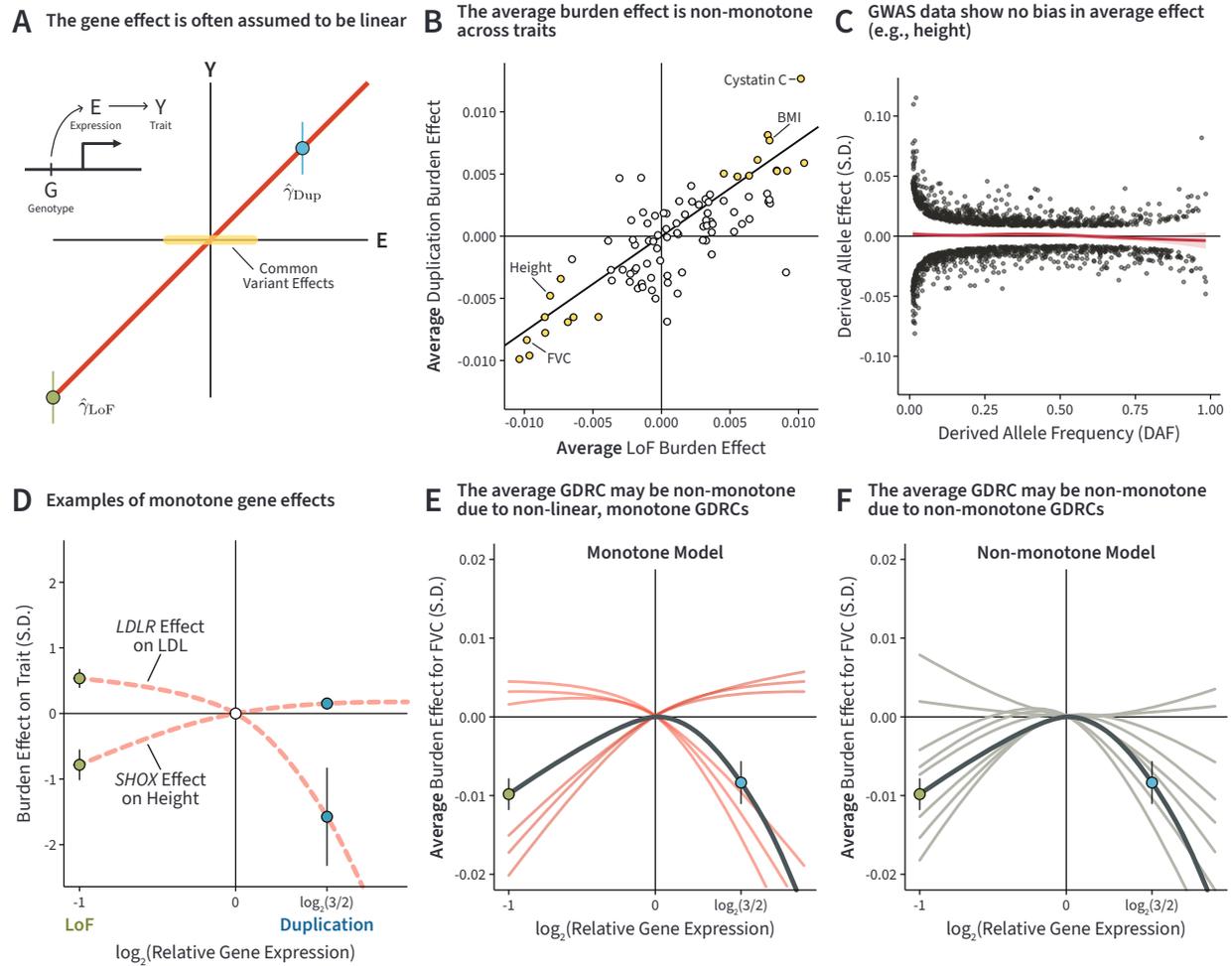


FIGURE 1. Explaining the paradoxical observations in burden data. **A.** The hypothetical relationship between gene dosage (E) and expected trait value (Y) is often assumed to be linear. LoF variants and duplications estimate the gene effect at different dosage values ($\hat{\gamma}_{\text{LoF}}$ and $\hat{\gamma}_{\text{Dup}}$). **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.

2 point on the GDRC is the hypothetical mean trait value of individuals whose gene expression is set to a
 3 particular level. In other words, the GDRC characterizes the relationship between expression and trait
 4 value for all possible dosage perturbations. The GDRC acts as a conceptual framework to unify various
 5 estimates of gene effects, allowing us to understand the source of non-linearity in the data.

6 Averaging over genetic backgrounds, a particular variant will have a specific effect on expression and
 7 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, $\hat{\gamma}_{\text{LoF}}$, is an estimate of a point
 9 on the left tail of the curve, while a burden estimate of the duplication effect, $\hat{\gamma}_{\text{Dup}}$, is an estimate of a

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

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

101 We suggest two models that could explain observations of non-monotone average GDRs. In the
102 *monotone model*, the non-monotone average GDR can be explained using only monotone gene-level
103 GDRs if the monotone curves are systematically buffered against one trait direction (Figure 1E). In the
104 *non-monotone model*, a subset of GDRs are non-monotone, and are more often non-monotone in a par-
105 ticular direction, driving the average GDR to be non-monotone (Figure 1F). These models are not mutu-
106 ally exclusive and may both contribute to the non-monotone average GDR observed for a trait.

Top hits from burden tests reveal both monotone and non-monotone gene effects

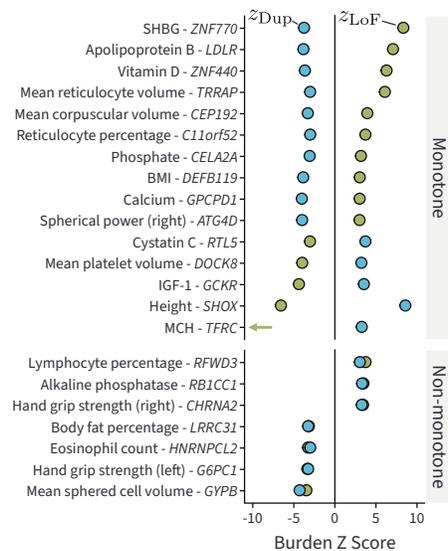


FIGURE 2. Top hits from burden tests. The Z scores from regression-based burden tests using LoF and duplication variants across various continuous traits. Top hits are separated into monotone and non-monotone relationships. The green arrow represents a Z score smaller than -10.

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

110 models could plausibly contribute to the observed average GDRCs.

111 MOST TRAITS ARE MONOTONE ON AVERAGE

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

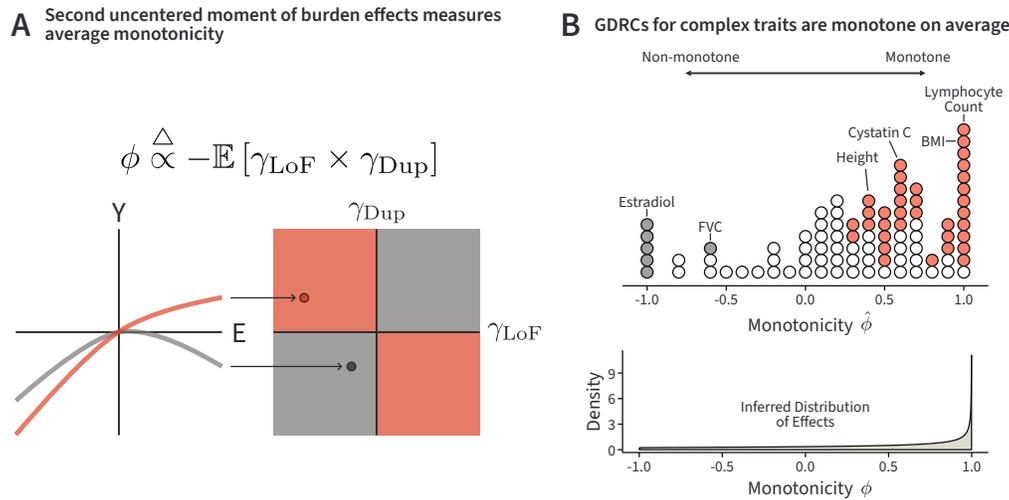


FIGURE 3. Monotonicity inference for traits. **A.** Monotone curves (red) and non-monotone curves (gray) fall in different quadrants on the burden effect plot. A natural estimate of comparing points in these quadrants is the second uncentered moment between burden effect sizes. **B.** Estimated values of $\hat{\phi}$. 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.

115 We started with the observation that monotone GDRCs imply that LoF variants and duplications
 116 should have effects in opposite directions (Figure 3A), which we visualized by plotting the LoF burden
 117 effect against the duplication burden effect. If a duplication and an LoF variant have a large effect in the
 118 same direction, the GDRC would be non-monotone and the product of their effects would be positive. If
 119 the GDRC is monotone, then LoF variants and duplications will have opposite effects, and their product
 120 will be negative. We define $\hat{\phi}$ to be a normalized version of the negative product of the effect sizes and
 121 term this the *monotonicity* (Methods, Appendix B). $\hat{\phi}$ is positive if genes with large effects on a trait tend
 122 to have monotone effects on average.

123 Across traits, we found that the burden signal was explained predominantly by monotone GDRCs, al-
 124 though five traits had significantly negative monotonicity estimates (Figures 3B and F.9). To get a sense
 125 of the average level of monotonicity across traits, we inferred a distribution over the monotonicity esti-
 126 mates across traits, which confirmed that both positive and negative $\hat{\phi}$ may be observed (Figure 3B). 71%
 127 of the density of the inferred distribution was over positive values of $\hat{\phi}$, while the rest was negative.

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

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

134 **NON-MONOTONE GENES ALIGN WITH THE DIRECTION OF AVERAGE EFFECTS**

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

141 The non-monotone model (Figure 1F) proposes that there are more non-monotone curves aligned
142 with the average burden effect than not. To assess this, we polarized ascertained gene-trait pairs based
143 on the direction of the average burden effect. Consistent with our model, non-monotone genes were
144 significantly more likely to be found in the quadrant concordant with the average effect (Figures 4A and
145 4B). As a specific example of a gene that falls in the concordant quadrant, the average LoF and duplication
146 burden effect is to reduce hand grip strength (Table F.1), and one of the top non-monotone genes, *G6PC1*,
147 reduces hand grip strength upon perturbation in either direction (Figure 2).

148 The presence of non-monotone GDRCs that contribute strongly to the average burden effect is sur-
149 prising. We can use a simple mechanistic model to see why such non-monotone GDRCs are unexpected.
150 Consider a gene that affects a trait via a single biological pathway (Figure 4C). Intuition about biological
151 mechanisms suggests that each step along a pathway should be monotone, but possibly non-linear. For
152 instance, protein levels are often buffered against large changes in gene expression [9], which results in a
153 non-linear, but monotone response of protein levels to gene expression (Figure 4C). Similarly, perturba-
154 tions of the expression levels of transcription factors can result in non-linear changes in the expression
155 of their targets [33, 34]. As the dosage effect percolates through a pathway towards the focal trait, these
156 curves compose with one another to form the relationship between gene dosage and trait. If each curve
157 along the pathway is monotone, then the overall relationship between gene expression and trait is math-
158 ematically guaranteed to also be monotone (Figure 4C, Appendix A).

159 However, non-monotone GDRCs can arise if a gene's effects flow through two or more pathways (Fig-
160 ure 4D). 3-hydroxybutyrate levels, which increase in blood during ketogenesis, provide an interpretable
161 example. *CYP11B1* has a non-monotone relationship with 3-hydroxybutyrate levels in the blood, with
162 both LoF variants and duplications increasing 3-hydroxybutyrate on average ($\hat{\gamma}_{\text{LoF}} = 0.15$, $z_{\text{LoF}} =$
163 3.03 , $\hat{\gamma}_{\text{Dup}} = 0.68$, $z_{\text{Dup}} = 2.66$). *CYP11B1* encodes a key adrenal enzyme involved in the production
164 of cortisol. Deficiency of the enzyme encoded by *CYP11B1*, referred to as 11 β -hydroxylase deficiency,
165 is a cause of a congenital adrenal condition that results in reduced cortisol levels and subsequent hy-
166 poglycemia [37]. Without sufficient glucose levels, the body generates increased amounts of ketones,

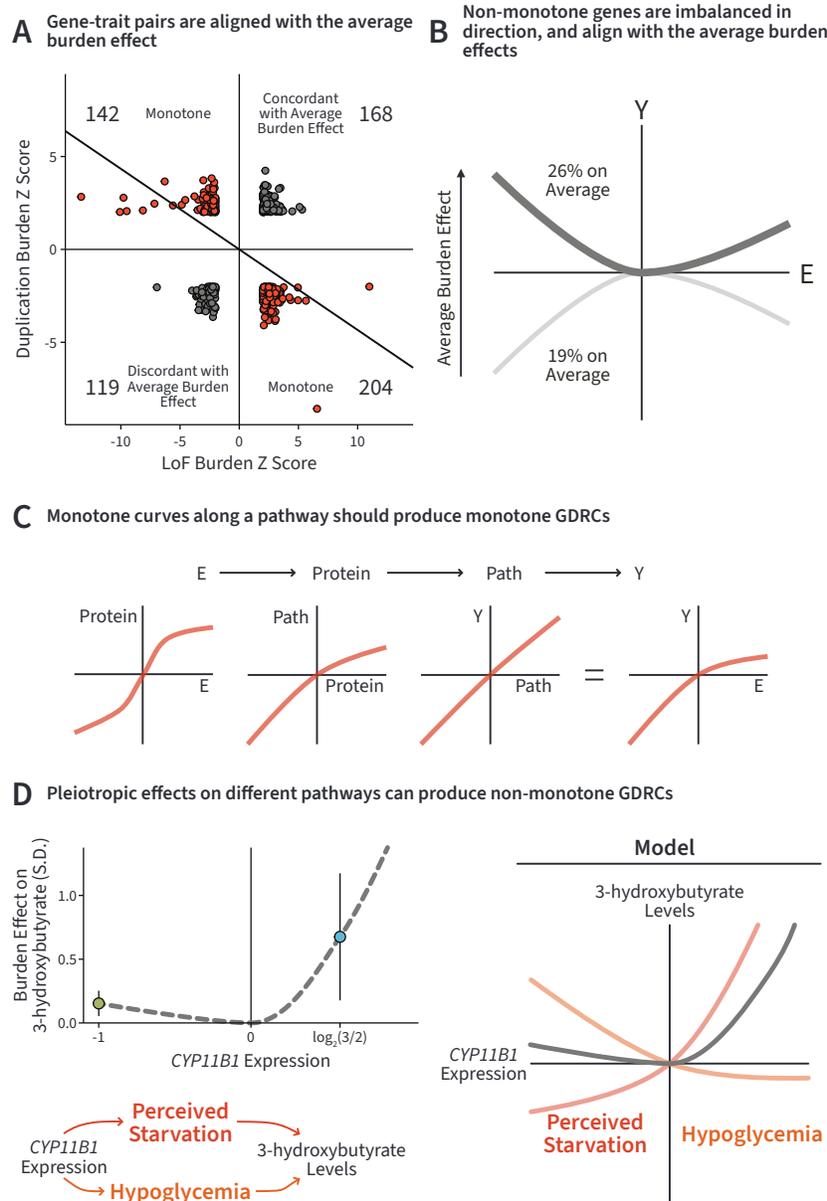


FIGURE 4. Large fraction of non-monotone GDRCs. **A.** Top gene-trait pairs ($|z| > 2$) for traits with significant average burden effects (Table F.1). The burden effects were polarized by the direction of the average burden effect. **B.** The proportion of concordant versus discordant non-monotone GDRCs 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.

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

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

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

176 **GENES ARE BUFFERED AGAINST ONE TRAIT DIRECTION**

177 The analysis of monotonicity and the imbalance in non-monotone GDRCs suggest that both the mono-
178 tone and non-monotone models are plausible explanations of the non-monotone average GDRCs. If
179 the mixture of GDRCs for the trait are depressed in one direction, then the average GDRC will be non-
1 0 monotone (Figure 5A). Another interpretation of these curves is that traits are more likely to be perturbed
1 1 in one direction than the other regardless of the direction of expression perturbation. Therefore, this is
1 2 a form of buffering that occurs at the trait level. We refer to this as *trait buffering*. We endeavored to
1 3 quantify how much each of our models contributed to the non-monotone average GDRC by measuring
1 4 and decomposing the trait buffering for various traits.

1 5 In the burden effect plot, curves experiencing trait buffering preferentially map to a region that is
1 6 away from a diagonal line (Figure 5A). The diagonal line represents perfectly linear GDRCs (Appendix E).
1 7 Points above the diagonal line correspond to GDRCs that are buffered against negative trait values, and
1 points below the diagonal line correspond to GDRCs buffered against positive trait values.

1 9 To develop some preliminary intuition about the trait buffering model, we estimated γ_{LoF} and γ_{Dup}
190 for each gene. Since burden effect estimates are noisy for individual genes, we performed Bayesian in-
191 ference using a flexible multivariate adaptive shrinkage (MASH) prior [42], which models the joint distri-
192 bution of γ_{LoF} and γ_{Dup} as a mixture of bivariate normal distributions. This prior distribution is learned
193 from the data. We used posterior samples of γ_{LoF} and γ_{Dup} for each gene to better understand the dis-
194 tribution of genes on the burden effect plot. As an example, the posterior means of γ_{LoF} and γ_{Dup} for
195 height are displayed in Figure 5B. Both monotone and non-monotone GDRCs are predominantly below
196 the diagonal, concordant with the direction of the average burden effect.

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

200 The sign of ξ indicates which direction is being buffered against. For example, our model predicts that
201 height should have a negative value of ξ and cystatin C should have a positive value of ξ . If GDRCs are not
202 preferentially buffered against either direction of a trait, our statistic should be close to zero. Indeed this
203 measure of trait buffering showed strong concordance with both the average LoF and duplication burden
204 effects (Figure 5C).

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

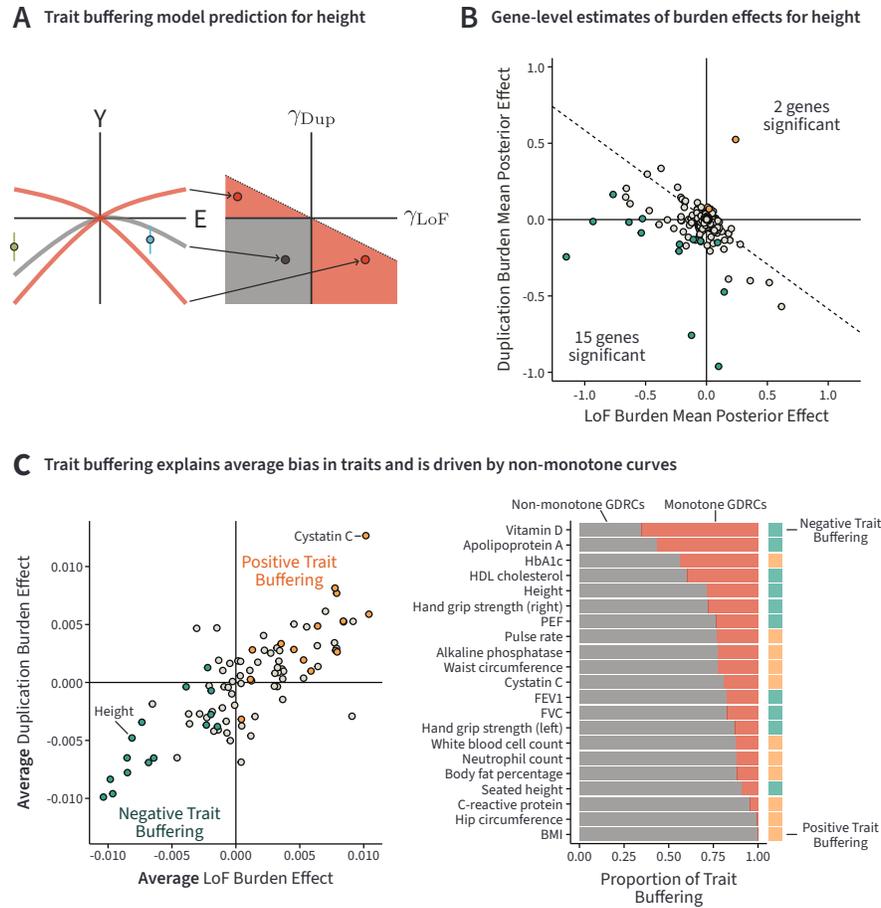


FIGURE 5. Trait buffering as a model for GDRC architecture. **A.** Combining both models of GDRCs contributing to the average burden effect. The GDRCs fall below a diagonal line for traits with negative average burden effects such as height. The average loss-of-function and duplication burden effects 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.

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

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

213 **A DYSREGULATION PHENOTYPE MAY BE UPSTREAM OF MANY COMPLEX TRAITS**

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

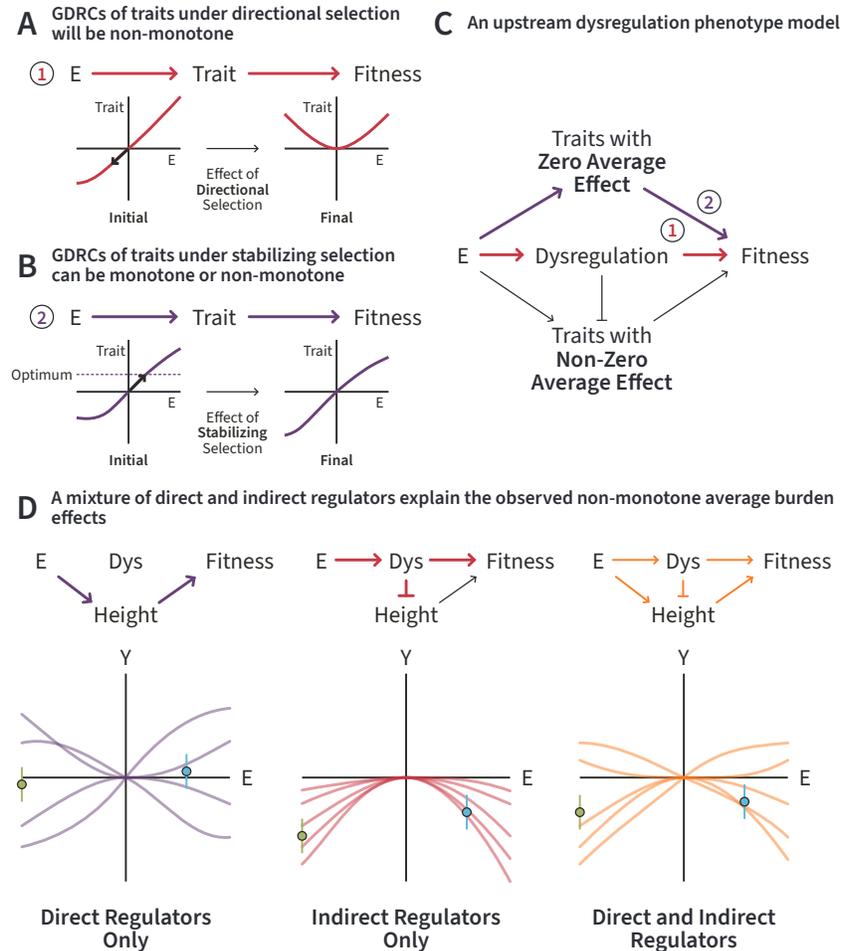


FIGURE 6. A theoretical model to explain trait buffering. **A.** GDRCs of traits under directional selection will become non-monotone over evolutionary time. **B.** GDRCs of traits under stabilizing selection will have no preference for GDRC shape. **C.** Traits with non-zero average burden effects have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

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

220 To see why these observations are signatures of directional selection, we consider how the average
 221 expression of a given gene will evolve over time in response to directional selection. Suppose a gene
 222 affects a trait that is under negative selection (Figure 6A). Then, selection will tune the expression of the
 223 gene until the trait is minimized, resulting in non-monotone GDRCs where both deletions and duplica-

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

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

230 Stabilizing selection affects GDRCs differently than directional selection. Suppose a gene affects a
231 trait under stabilizing selection (Figure 6B). Stabilizing selection acts to maintain an optimum trait value,
232 and will tune the expression of the gene to reach the optimum trait value. Unlike in the case of directional
233 selection, the local shape of the GDRC around the optimum trait value is not restricted to minima. In this
234 case, we do not expect any particular GDRC shape to dominate, and we do not expect LoF variants or
235 duplications to be biased in a particular direction.

236 To explain the incongruous non-zero average burden effects and zero average GWAS effect, we pro-
237 pose that traits with non-monotone average GDRCs are downstream of an unobserved trait under neg-
238 ative selection that we refer to as ‘dysregulation’ (Figure 6C). If the focal trait is downstream of dysregu-
239 lation, the GDRC of the gene for the focal trait may be shaped by both forms of selection depending on
240 if the gene is a direct or indirect regulator of the focal trait (Figure 6D). For genes that act primarily on
241 the trait via dysregulation, the GDRCs will be non-monotone and in the direction of increasing dysregu-
242 lation. For genes that directly modulate the focal trait, we expect no particular preference for monotone
243 or non-monotone GDRCs. Genes that act on the focal trait both directly and via dysregulation explain the
244 monotone curves that experience trait buffering.

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

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

258 We believe that this unobserved latent phenotype is a measure of dysregulation because the down-
259 stream traits with large non-monotone average burden effects are readouts of general dysfunction. The

260 traits with significant non-monotone average effects (Table F.1) include measures of organ dysfunction
261 such as alkaline phosphatase, cystatin C, and C-reactive protein, which all have positive average LoF and
262 duplication effects. Put another way, both deletion or duplication of a random gene tend to increase dis-
263 ease risk and measures of organ dysfunction. Similarly, measures of lung capacity and muscle strength
264 have negative average LoF and duplication effects. Overall, the average effect is often in the direction of
265 dysregulation across the continuous traits we studied.

266 Recent analyses of predicted biological age in the UKB also support the presence of a dysregulation
267 phenotype [48, 49]. Age is associated with general dysregulation and disease [50]. Strong predictors of
268 biological age in predictive models include FEV1, cystatin C, HbA1c, alkaline phosphatase, and hand grip
269 strength [48, 49], suggesting that these continuous traits with large non-monotone average effects may
270 act as readouts of dysregulation.

271 **DISCUSSION**

272 In this study, we interpreted dosage-perturbing rare variants through the lens of the GDRC. Our analysis
273 suggests that both monotone and non-monotone GDRCs contribute to the concordant average burden
274 effects of LoF and duplication variants. This results in GDRCs that are buffered against one trait direc-
275 tion. We hypothesized that this may be explained by the effect of genes on an upstream dysregulation
276 phenotype. The shape of GDRCs may thus be molded by the effects of directional selection.

277 **IMPLICATIONS OF THE OBSERVED CHARACTERISTICS OF GENE DOSAGE RESPONSE CURVES**

278 In a “linear world” where dosage has a linear relationship with trait, ascertaining any point on the GDRC
279 along the dosage spectrum would carry all the information about the gene’s effect on trait. However,
280 biology is inherently non-linear, and our analysis shows that these assumptions break down for variants
281 with large effects. This has significant implications for drug design. If the GDRC of the target gene is
282 non-monotone, perturbation in either direction may only increase disease risk. This occasional lack of
283 concordance between LoF and duplication burden effects motivates the use of assays that test both ends
284 of the dosage spectrum [51]. The shape of the GDRC implied by the burden estimates can also be useful
285 prior for fine-mapping causal variants [52] and linking variants to genes [53–57].

286 The assumption of linearity in TWAS [5, 6, 58] and Mendelian randomization [59] approaches may
287 cause estimated effect sizes to be close to zero for non-monotone genes. The ability of TWAS to recover
288 trait-associated genes with non-linear GDRCs needs to be explored further. Yet, TWAS-type methods
289 would be invaluable for understanding the behavior of the GDRC in the physiological range of expres-
290 sion, thus motivating the continued development of such methods.

291 **TOWARDS THE INFERENCE OF GENE-LEVEL CURVES ACROSS TISSUES AND CONTEXTS**

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

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

301 Even with methodological advances, it is unlikely that we will be able to infer all points of the GDRC
302 of a gene with population-level data alone. The effects of selection make it challenging to detect variants
303 in genes with large effects on traits [60, 61]. This suggests that the ability to effectively infer the GDRC will
304 vary across the genome, and likely be most challenging for genes that are particularly important to the
305 biology of traits.

306 These limitations of population-level data suggest that sequence-to-expression models [62–65] and
307 experimental approaches [66–68] will play a critical role in fleshing out the GDRC for key genes. Missense
308 variants, which in aggregate have a measurable effect on complex traits [30], are challenging to integrate
309 into the GDRC framework. Accurate computational effect prediction [69] or experimental approaches
310 such as deep mutational scanning [70] are critical for translating the effect of missense variants into an
311 “effective dosage”, which would allow them to be placed on the GDRC. Similarly, for common variants, we
312 are restricted to using dosage effect estimates from eQTL studies in tissues that have been assayed, which
313 are often post-mortem and from adult donors [71]. However, we cannot feasibly assay all tissues and cell
314 types in all contexts. Thus, predicting common variant effects from sequence and multi-omic context
315 or using experimental approaches such as massively parallel reporter assays [67] will be necessary to
316 estimate full GDRCs.

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

322 **METHODS**

323 The goal of our data analysis approach was to estimate burden summary statistics for LoF variants, whole-
324 gene deletions, whole-gene duplications, and partial deletions using a consistent set of decisions and
325 tools. This consistency between the burden tests allowed us to compare effect sizes across the variant
326 classes. We performed burden test using synonymous variants as a negative control to estimate any
327 residual confounding, which we detail in Appendix F. We improved upon previous LoF burden approaches
328 by using variable frequency filters based on gene constraint (Appendix F). We also improved upon prior
329 duplication burden tests in the UKB [23–25, 27] by using copy number calls from WGS data rather than
330 microarray data. CNV calls from WGS have the advantage of more accurate copy number estimation,
331 higher resolution of CNV breakpoints, and better detection of rare CNVs [72].

332 To make our approach amenable to researchers without access to individual level data, we modeled
333 the summary statistics rather than the underlying genotype data. We use the regression with summary
334 statistics (RSS) likelihood [73], which provides a likelihood model for association test summary statistics.
335 To our knowledge, this is the first use of the RSS likelihood for burden regression. To demonstrate its
336 applicability and correctness, we performed extensive simulations, which are catalogued in Appendix D.

337 **INDIVIDUALS**

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

344 **PHENOTYPES**

345 We curated 410 continuous phenotypes in the UKB on which to perform burden tests, of which 91 were
346 used for the main analysis in the paper. We required phenotypes to have measurements in at least 40K
347 individuals. We also excluded any traits where more than 50% of the instances match the mode of the
348 trait. This is to remove quasi-categorical traits, and is more relaxed compared to Backman et al., who
349 exclude traits with a cutoff of 20%. For phenotypes that were collected more than once in returning visits,
350 we use the first instance of the observation for each individual. Backman et al. took the mean across such
351 instances, but we believe that this would reduce the noise for some individuals more than others in a not
352 completely at random fashion. It would also make it challenging to interpret the relationship between
353 mean phenotype and covariates that are observed at specific instances, such as age. Some of our traits
354 had arrayed measurements, but they were all taken in the same visit so we decided to use the mean value
355 across the arrayed items.

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

35 GENOTYPES

359 We included 15 genotyping principal components (PCs), estimated previously in the UKB [75]. Backman
360 et al. performed ancestry-specific analyses and included 10 genotyping PCs for each ancestry. Since we
361 used all 460K WES samples, we included an additional 5 PCs to address potential stratification. Our anal-
362 ysis of synonymous variants (Appendix F) supported this choice.

363 Following the lead of Backman et al., we included 20 rare variant PCs, by which we mean PCs derived
364 from rare WES variants. We subsampled 300K variants uniformly at random from the rare fraction, which
365 we defined as variants with minor allele count (MAC) greater than 20 and minor allele frequency (MAF)
366 less than 1%. We used an approximate principal components analysis (PCA) algorithm implemented in
367 FlashPCA2 [76].

368 We restricted our burden analysis to LoF sites with MAF less than 1%. High-confidence LoF sites were
369 annotated using Ensembl's Variant Effect Predictor [77] with the LOFTEE plugin [78]. LOFTEE filters re-
370 move annotated LoF sites that might escape nonsense-mediated decay (NMD). Instead of using progres-
371 sively stricter MAF cutoffs as in Backman et al., we use all LoF sites with MAF less than 1% but require that
372 the misannotation probability be less than 10%. The misannotation probability [61] takes into account
373 both the frequency of the site and the constraint experienced by the gene to determine a posterior prob-
374 ability of being misannotated as a LoF. These probabilities were previously reported for LoF-introducing
375 single nucleotide polymorphisms (SNPs) for genes with estimated s_{het} values [61]. A small number of
376 genes do not have s_{het} estimates, and not all variants in the WES data were SNPs. For genes with missing
377 s_{het} values, s_{het} was imputed using the mean s_{het} value. Then, for all LoF variants missing a misannota-
378 tion probability, misannotation probabilities were imputed using k-nearest neighbors (kNN) regression
379 with 10 nearest neighbors as determined by allele frequency and s_{het} . A 10% cutoff on this probability
380 retains 96.46% of variants. We demonstrated that relative to standard frequency cutoffs, using a misan-
381 notation probability cutoff increases the signal in the burden tests without increasing the size of standard
382 errors (Appendix F).

383 CNVs were previously called in the UKB using Illumina's DRAGEN software [31, 79]. A deletion was de-
384 fined as any loss that affected the entire gene body with an additional 1 Kbp flanking region. A duplication
385 was defined as any gain that affected the entire gene body with an additional 1 Kbp flanking region. Any
386 deletions that affected only part of the gene body were categorized as potential loss-of-function (pLoF)
387 alleles. For deletions and pLoF variants, the alternative allele coded the number of copy loss events. For
388 duplications, a copy number of two was coded as the reference homozygote and a copy number greater
389 than two was coded as a heterozygote. We observed that some samples had a large amount of genome-
390 wide copy gain or loss, which is likely a genotyping error (Appendix F). To account for this, we removed
391 any sample with greater than 10 Mbp of deleted or duplicated genome sequence.

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

394 BURDEN TESTS

395 We used REGENIE to perform gene-level burden tests [32]. For the whole-genome regression, which is the
396 first step in REGENIE, SNPs from the genotyping array were pruned with a 1000 variant sliding window
397 with 100 variant shifts and an R^2 threshold of 0.9. SNPs were then filtered to have MAF > 1%, genotype
39 missingness < 10%, and Hardy-Weinberg equilibrium (HWE) test p-value < 10^{-15} .

399 For LoF burden tests, we used the following covariates: age, sex, age-by-sex, age squared, 15 genotyping
400 PCs, 20 rare variant PCs, and WES batch. For the duplication, deletion, or pLoF burden tests, we
401 used the following covariates: age, sex, age-by-sex, age squared, 15 genotyping PCs, 20 rare variant PCs,
402 and WGS batch. When analyzing NMR metabolites, we additionally included the spectrometer and pro-
403 cessing batch as covariates.

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

407 MONOTONICITY MODEL

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

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

409 where \mathbf{X} and \mathbf{Z} are the burden genotypes from LoF variants and duplications respectively. The error $\boldsymbol{\epsilon}$
410 was assumed to be drawn from a normal distribution. Polygenic signal, population stratification, and
411 other sources of confounding were accounted for when performing the burden tests by including covari-
412 ates. We built a hierarchical model by specifying a distribution on the latent effect sizes of the j th gene:

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

413 This allowed us to model the quantities of interest. The average effect of deletions and duplications is
414 represented by $\bar{\gamma}_1 \mathbf{1}$ and $\bar{\gamma}_2 \mathbf{1}$ respectively. We defined the monotonicity, $\phi \in [-1, 1]$, to be a quantity that
415 is proportional to the negative uncentered second moment of the latent effect sizes,

$$\phi \stackrel{\Delta}{\propto} -[\gamma_{1j}\gamma_{2j}].$$

416 Thus, if genes with large effects tend to have opposite signs on average, ϕ will be closer to 1. Similarly, if
417 genes with large effects tend to have the same sign on average, ϕ will be closer to -1.

41 The observed burden effect sizes for the j th gene, $\hat{\gamma}_{1j}$ and $\hat{\gamma}_{2j}$, are noisy estimators of the underlying
419 latent effect size. Furthermore, due to correlation between duplication burden genotypes, $\hat{\gamma}_{2j}$ is not an
420 unbiased estimator of γ_{2j} because the observed effect size is inflated by the effect of nearby genes. To
421 account for these concerns, we utilized the RSS likelihood [73]. An extensive description of the model is
422 provided in Appendix B.

423 We developed unbiased estimators of $\bar{\gamma}_1$ and $\bar{\gamma}_2$. Since the monotonicity is a complex function of
424 the parameters of the prior, it was challenging to derive unbiased estimators. Instead, we opted to use
425 maximum likelihood estimation for ϕ , which guarantees consistency. Inference and simulations under
426 this model are described in Appendices C and D respectively.

427 **BUFFERING MODEL**

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

432 To address this, we modified our monotonicity model to use the MASH model prior [42], which is a
433 flexible multivariate prior. We continued to use the RSS likelihood to model the observed effect sizes.
434 MASH uses a mixture of K multivariate normal distributions to model the latent space:

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

435 Since the true latent distribution is unknown, MASH uses a grid of fixed covariance matrices and learns
436 the mixture weights π using empirical Bayes. In its original implementation, MASH assumes indepen-
437 dence between observed samples, allowing for efficient optimization. Since the RSS likelihood neces-
438 sarily induces dependencies between the observations via linkage disequilibrium (LD), we implemented
439 a stochastic approximation expectation maximization (SAEM) algorithm to optimize π [81, 82]. An alter-
440 native approach to this model has been proposed that uses variational inference [83–85].

441 As we defined it, the amount of directional buffering can be estimated by taking the dot product of the
442 latent effect size with the normal vector of the hyperplane separating the two types of buffering, which
443 we explain further in Appendix E. We used the posterior distribution over γ to estimate

$$\xi \triangleq \left[\log_2 \left(\frac{3}{2} \right) \gamma_{1j} + \gamma_{2j} \mid \hat{\gamma} \right] ,$$

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

446 GDRC:

$$\begin{aligned}\xi_{\text{nm}} &= \left[\log_2 \left(\frac{3}{2} \right) \gamma_{1j} + \gamma_{2j} \mid \hat{\gamma}_{1j} \gamma_{2j} > 0 \right] \\ \xi_{\text{m}} &= \left[\log_2 \left(\frac{3}{2} \right) \gamma_{1j} + \gamma_{2j} \mid \hat{\gamma}_{1j} \gamma_{2j} < 0 \right].\end{aligned}$$

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

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

449 ξ is zero under the null model where GDRCs are equally likely to be buffered against increasing or de-
450 creasing the trait. The details of the model can be found in Appendix E.

451 DATA ANALYSIS

452 Throughout, we use unbiased maximum likelihood estimates (MLEs) for the duplication burden esti-
453 mates when they are displayed in the paper, which account for LD with neighboring genes.

454 In Figure 1B, we used total least squares to fit a regression line with no intercept. Weights in the
455 regression corresponded to the inverse variance of the estimates. To provide a better fit, we used multiple
456 initial parameter values. We retrieved GWAS summary statistics from the UKB to generate trumpet plots
457 (Figures 1C and F.8). The collection and analysis of these data is described in Appendix F. We fit a locally
45 estimated scatterplot smoothing (LOESS) curve to visualize the conditional mean of the derived allele
459 effect given the derived allele frequency.

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

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

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

473 We estimated confident separation from the diagonal line in Figure 5B using the local false sign rate
474 [86]. This rate is calculated using the number of posterior samples that fall on either side of the diagonal

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

$$\frac{\xi_{nm} \mathbb{P}(\gamma_{1j} \gamma_{2j} > 0)}{\xi_m \mathbb{P}(\gamma_{1j} \gamma_{2j} < 0)} .$$

479 DATA AVAILABILITY

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

486 CODE AVAILABILITY

487 Scripts used to process and analyze data are deposited in Zenodo at [https://doi.org/10.5281/zenodo.1](https://doi.org/10.5281/zenodo.13864152)
488 3864152. Scripts were executed either on the UK Biobank Research Analysis Platform ([https://ukbiob](https://ukbiobank.dnanexus.com/)
489 [ank.dnanexus.com/](https://ukbiobank.dnanexus.com/)) or on the Sherlock High-Performance Computing Cluster managed by the Stanford
490 Research Computing Center (<https://www.sherlock.stanford.edu/>).

491 ACKNOWLEDGMENTS

492 We would like to thank the members of the Pritchard lab for their feedback on this project and manuscript.
493 In addition, we thank Yi Ding and Zeyun Lu (Gusev Lab) for reading over and providing feedback on the
494 initial draft. This research has been conducted using data from the UK Biobank ([https://www.ukbiob](https://www.ukbiobank.ac.uk/)
495 [ank.ac.uk/](https://www.ukbiobank.ac.uk/)), a major biomedical database. We would like to thank all participants and researchers in-
496 volved in the UK Biobank for their contribution, and Nightingale Health PLC for early access to the NMR
497 metabolomics data. Computing for this project was performed on the Sherlock High-Performance Com-
498 puting Cluster. We would like to thank Stanford University and the Stanford Research Computing Center
499 for providing computational resources and support that contributed to this research. N.M. was supported
500 by a National Science Foundation Graduate Research Fellowship and Stanford's Knight-Hennessy Schol-
501 ars Program. C.J.S. was supported by Stanford's Knight-Hennessy Scholars Program and the Stanford
502 Center for Computational, Evolutionary and Human Genomics. H.Z. was supported by Stanford Biology
503 Department's graduate student assistantship. This work was supported by the National Institutes of
504 Health (R01HG011432 and R01HG008140 to J.K.P.).

505 AUTHOR CONTRIBUTIONS

506 **N.M.** Methodology, Software, Formal Analysis, Data Curation, Writing – Original Draft, Visualization.
507 **C.J.S.** Methodology, Software, Writing – Original Draft, Visualization. **H.Z.** Methodology, Software, Writing
508 – Original Draft, Visualization. **J.P.S.** Conceptualization, Methodology, Writing – Original Draft, Supervision.
509 **J.K.P.** Conceptualization, Methodology, Writing – Original Draft, Supervision, Project Administration, Funding Acquisition.
510

511 REFERENCES

- 512 [1] Melina Claussnitzer et al. “A brief history of human disease genetics”. *Nature* 577.7789 (Jan. 2020), pp. 179–
513 189. doi: 10.1038/s41586-019-1879-7.
- 514 [2] Elliot Sollis et al. “The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource”. *Nucleic Acids
515 Research* 51.D1 (Jan. 2023), pp. D977–D985. doi: 10.1093/nar/gkac1010.
- 516 [3] Matthew T. Maurano et al. “Systematic Localization of Common Disease-Associated Variation in Regulatory
517 DNA”. *Science* 337.6099 (Sept. 2012), pp. 1190–1195. doi: 10.1126/science.1222794.
- 518 [4] Anshul Kundaje et al. “Integrative analysis of 111 reference human epigenomes”. *Nature* 518.7539 (Feb. 2015),
519 pp. 317–330. doi: 10.1038/nature14248.
- 520 [5] Eric R. Gamazon et al. “A gene-based association method for mapping traits using reference transcriptome
521 data”. *Nature Genetics* 47.9 (Aug. 2015), pp. 1091–1098. doi: 10.1038/ng.3367.
- 522 [6] Alexander Gusev et al. “Integrative approaches for large-scale transcriptome-wide association studies”. *Nature
523 Genetics* 48.3 (Feb. 2016), pp. 245–252. doi: 10.1038/ng.3506.
- 524 [7] Farhad Hormozdiari et al. “Colocalization of GWAS and eQTL Signals Detects Target Genes”. *The American
525 Journal of Human Genetics* 99.6 (Dec. 2016), pp. 1245–1260. doi: 10.1016/j.ajhg.2016.10.003.
- 526 [8] Wouter Meuleman et al. “Index and biological spectrum of human DNase I hypersensitive sites”. *Nature
527* 584.7820 (July 2020), pp. 244–251. doi: 10.1038/s41586-020-2559-3.
- 528 [9] Silvia Stingele et al. “Global analysis of genome, transcriptome and proteome reveals the response to aneu-
529 ploidy in human cells”. *Molecular Systems Biology* 8.1 (Sept. 2012), p. 608. doi: 10.1038/msb.2012.40.
- 530 [10] Terry Hassold and Patricia Hunt. “To err (meiotically) is human: the genesis of human aneuploidy”. *Nature
531 Reviews Genetics* 2.4 (Apr. 2001), pp. 280–291. doi: 10.1038/35066065.
- 532 [11] C D Lehman et al. “Trisomy 13 syndrome: prenatal US findings in a review of 33 cases.” *Radiology* 194.1 (Jan.
533 1995), pp. 217–222. doi: 10.1148/radiology.194.1.7997556.
- 534 [12] Anna Cereda and John C. Carey. “The trisomy 18 syndrome”. *Orphanet Journal of Rare Diseases* 7.1 (Oct.
535 2012), p. 81. doi: 10.1186/1750-1172-7-81.
- 536 [13] Marilyn J. Bull. “Down Syndrome”. *New England Journal of Medicine* 382.24 (June 2020), pp. 2344–2352. doi:
537 10.1056/NEJMra1706537.
- 538 [14] Xin Shao et al. “Copy number variation is highly correlated with differential gene expression: a pan-cancer
539 study”. *BMC medical genetics* 20.1 (Nov. 2019), p. 175. doi: 10.1186/s12881-019-0909-5.
- 540 [15] Peter H. Sudmant et al. “Global diversity, population stratification, and selection of human copy-number
541 variation”. *Science* 349.6253 (Sept. 2015), aab3761. doi: 10.1126/science.aab3761.
- 542 [16] Haley J. Abel et al. “Mapping and characterization of structural variation in 17,795 human genomes”. *Nature
543* 583.7814 (May 2020), pp. 83–89. doi: 10.1038/s41586-020-2371-0.
- 544 [17] Ryan L. Collins et al. “A structural variation reference for medical and population genetics”. *Nature* 581.7809
545 (May 2020), pp. 444–451. doi: 10.1038/s41586-020-2287-8.

- 546 [18] Mohamed A. Almarri et al. “Population Structure, Stratification, and Introgression of Human Structural Vari-
547 ation”. *Cell* 182.1 (July 2020), 189–199.e15. doi: 10.1016/j.cell.2020.05.024.
- 54 [19] Andrew Dauber et al. “Genome-wide Association of Copy-Number Variation Reveals an Association between
549 Short Stature and the Presence of Low-Frequency Genomic Deletions”. *The American Journal of Human Ge-
550 netics* 89.6 (Dec. 2011), pp. 751–759. doi: 10.1016/j.ajhg.2011.10.014.
- 551 [20] Eleanor Wheeler et al. “Genome-wide SNP and CNV analysis identifies common and low-frequency variants
552 associated with severe early-onset obesity”. *Nature Genetics* 45.5 (Apr. 2013), pp. 513–517. doi: 10.1038/ng.
553 2607.
- 554 [21] Aurélien Macé et al. “CNV-association meta-analysis in 191,161 European adults reveals new loci associated
555 with anthropometric traits”. *Nature Communications* 8.1 (Sept. 2017), p. 744. doi: 10.1038/s41467-017-
556 00556-x.
- 557 [22] Christian R. Marshall et al. “Contribution of copy number variants to schizophrenia from a genome-wide
55 study of 41,321 subjects”. *Nature Genetics* 49.1 (Nov. 2016), pp. 27–35. doi: 10.1038/ng.3725.
- 559 [23] Matthew Aguirre, Manuel A. Rivas, and James Priest. “Phenome-wide Burden of Copy-Number Variation in
560 the UK Biobank”. *The American Journal of Human Genetics* 105.2 (Aug. 2019), pp. 373–383. doi: 10.1016/j.
561 ajhg.2019.07.001.
- 562 [24] Chiara Auwerx et al. “The individual and global impact of copy-number variants on complex human traits”.
563 *The American Journal of Human Genetics* 109.4 (Apr. 2022), pp. 647–668. doi: 10.1016/j.ajhg.2022.02.010.
- 564 [25] Margaux L. A. Hujoel et al. “Influences of rare copy-number variation on human complex traits”. *Cell* 185.22
565 (Oct. 2022), 4233–4248.e27. doi: 10.1016/j.cell.2022.09.028.
- 566 [26] Ryan L. Collins et al. “A cross-disorder dosage sensitivity map of the human genome”. *Cell* 185.16 (Aug. 2022),
567 3041–3055.e25. doi: 10.1016/j.cell.2022.06.036.
- 56 [27] Chiara Auwerx et al. *Copy-number variants as modulators of common disease susceptibility*. Aug. 2023. doi:
569 10.1101/2023.07.31.23293408.
- 570 [28] Chia-Yen Chen et al. “The impact of rare protein coding genetic variation on adult cognitive function”. *Nature
571 Genetics* 55.6 (June 2023), pp. 927–938. doi: 10.1038/s41588-023-01398-8.
- 572 [29] Yunfeng Huang et al. “Rare genetic variants impact muscle strength”. *Nature Communications* 14.1 (June
573 2023), p. 3449. doi: 10.1038/s41467-023-39247-1.
- 574 [30] Joshua D. Backman et al. “Exome sequencing and analysis of 454,787 UK Biobank participants”. *Nature
575* 599.7886 (Oct. 2021), pp. 628–634. doi: 10.1038/s41586-021-04103-z.
- 576 [31] Shuwei Li et al. *Whole-genome sequencing of half-a-million UK Biobank participants*. Dec. 2023. doi: 10.1101/
577 2023.12.06.23299426.
- 57 [32] Joelle Mbatchou et al. “Computationally efficient whole-genome regression for quantitative and binary
579 traits”. *Nature Genetics* 53.7 (May 2021), pp. 1097–1103. doi: 10.1038/s41588-021-00870-7.
- 58 [33] Sahin Naqvi et al. “Precise modulation of transcription factor levels identifies features underlying dosage
581 sensitivity”. *Nature Genetics* 55.5 (May 2023), pp. 841–851. doi: 10.1038/s41588-023-01366-2.
- 582 [34] Júlia Domingo et al. *Non-linear transcriptional responses to gradual modulation of transcription factor
583 dosage*. Mar. 2024. doi: 10.1101/2024.03.01.582837.
- 584 [35] Sébastien Jacquemont et al. “Mirror extreme BMI phenotypes associated with gene dosage at the chromo-
585 some 16p11.2 locus”. *Nature* 478.7367 (Aug. 2011), pp. 97–102. doi: 10.1038/nature10406.
- 586 [36] Christelle Golzio et al. “KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2
587 copy number variant”. *Nature* 485.7398 (May 2012), pp. 363–367. doi: 10.1038/nature11091.
- 58 [37] Tülay Güran. “Latest Insights on the Etiology and Management of Primary Adrenal Insufficiency in Children”.
59 *J Clin Res Pediatr Endocrinol* 9.60 (2017), pp. 9–22. doi: 10.4274/jcrpe.2017.S002.

- 590 [38] Caleb B. Cantrell and Shamim S. Mohiuddin. *Biochemistry, Ketone Metabolism*. StatPearls Publishing, Treas-
591 sure Island (FL), Mar. 2020.
- 592 [39] Agata Juszcak, Damian Morris, and Ashley Grossman. “Cushing’s Syndrome”. *Endotext*. Ed. by Kenneth R.
593 Feingold et al. South Dartmouth (MA): MDText.com, Inc., 2000.
- 594 [40] Juliette Emmerich et al. “Cushing’s Syndrome: Development of Highly Potent and Selective CYP11B1 In-
595 hibitors of the (Pyridylmethyl)pyridine Type”. *Journal of Medicinal Chemistry* 56.15 (Aug. 2013), pp. 6022–
596 6032. doi: 10.1021/jm400240r.
- 597 [41] Vladimir Stojanovic and Sherri Ihle. “Role of beta-hydroxybutyric acid in diabetic ketoacidosis: A review”. *The*
59 *Canadian Veterinary Journal* 52.4 (Apr. 2011), pp. 426–430.
- 599 [42] Sarah M. Urbut et al. “Flexible statistical methods for estimating and testing effects in genomic studies with
600 multiple conditions”. *Nature Genetics* 51.1 (Nov. 2018), pp. 187–195. doi: 10.1038/s41588-018-0268-8.
- 601 [43] Yuval B. Simons et al. “A population genetic interpretation of GWAS findings for human quantitative traits”.
602 *PLOS Biology* 16.3 (Mar. 2018), e2002985. doi: 10.1371/journal.pbio.2002985.
- 603 [44] Guy Sella and Nicholas H. Barton. “Thinking About the Evolution of Complex Traits in the Era of Genome-
604 Wide Association Studies”. *Annual Review of Genomics and Human Genetics* 20.1 (Aug. 2019), pp. 461–493.
605 doi: 10.1146/annurev-genom-083115-022316.
- 606 [45] Yuval B. Simons et al. *Simple scaling laws control the genetic architectures of human complex traits*. Oct. 2022.
607 doi: 10.1101/2022.10.04.509926.
- 60 [46] E. Koch et al. *Genetic association data are broadly consistent with stabilizing selection shaping human com-
609 mon diseases and traits*. July 2024. doi: 10.1101/2024.06.19.599789.
- 610 [47] Roshni A. Patel et al. *Conditional frequency spectra as a tool for studying selection on complex traits in
611 biobanks*. June 2024. doi: 10.1101/2024.06.15.599126.
- 612 [48] Mei Sum Chan et al. “A Biomarker-based Biological Age in UK Biobank: Composition and Prediction of Mor-
613 tality and Hospital Admissions”. *The Journals of Gerontology: Series A* 76.7 (July 2021), pp. 1295–1302. doi:
614 10.1093/gerona/glab069.
- 615 [49] Sergiy Libert, Alex Chekholko, and Cynthia Kenyon. *A mathematical model that predicts human biological
616 age from physiological traits identifies environmental and genetic factors that influence aging*. Oct. 2023. doi:
617 10.1101/2022.04.14.488358.
- 61 [50] Carlos López-Otín et al. “The Hallmarks of Aging”. *Cell* 153.6 (June 2013), pp. 1194–1217. doi: 10.1016/j.cell.
619 2013.05.039.
- 620 [51] Ralf Schmidt et al. “CRISPR activation and interference screens decode stimulation responses in primary
621 human T cells”. *Science* 375.6580 (Feb. 2022), eabj4008. doi: 10.1126/science.abj4008.
- 622 [52] Anna Hutchinson, Jennifer Asimit, and Chris Wallace. “Fine-mapping genetic associations”. *Human Molecular
623 Genetics* 29.R1 (Sept. 2020), R81–R88. doi: 10.1093/hmg/ddaa148.
- 624 [53] Tuuli Lappalainen and Daniel G. MacArthur. “From variant to function in human disease genetics”. *Science*
625 373.6562 (Sept. 2021), pp. 1464–1468. doi: 10.1126/science.abi8207.
- 626 [54] Joseph Nasser et al. “Genome-wide enhancer maps link risk variants to disease genes”. *Nature* 593.7858 (Apr.
627 2021), pp. 238–243. doi: 10.1038/s41586-021-03446-x.
- 62 [55] Noah J Connally et al. “The missing link between genetic association and regulatory function”. *eLife* 11 (Dec.
629 2022). Ed. by Jonathan Flint and Molly Przeworski, e74970. doi: 10.7554/eLife.74970.
- 630 [56] Steven Gazal et al. “Combining SNP-to-gene linking strategies to identify disease genes and assess disease
631 omnigenicity”. *Nature Genetics* 54.6 (June 2022), pp. 827–836. doi: 10.1038/s41588-022-01087-y.
- 632 [57] Andreas R. Gschwind et al. *An encyclopedia of enhancer-gene regulatory interactions in the human genome*.
633 Nov. 2023. doi: 10.1101/2023.11.09.563812.

- 634 [58] Nicholas Mancuso et al. “Probabilistic fine-mapping of transcriptome-wide association studies”. *Nature Ge-*
635 *netics* 51.4 (Mar. 2019), pp. 675–682. doi: 10.1038/s41588-019-0367-1.
- 636 [59] Zhihong Zhu et al. “Integration of summary data from GWAS and eQTL studies predicts complex trait gene
637 targets”. *Nature Genetics* 48.5 (Mar. 2016), pp. 481–487. doi: 10.1038/ng.3538.
- 63 [60] Hakhamanesh Mostafavi et al. “Systematic differences in discovery of genetic effects on gene expression and
639 complex traits”. *Nature Genetics* (Oct. 2023), pp. 1–10. doi: 10.1038/s41588-023-01529-1.
- 640 [61] Tony Zeng et al. “Bayesian estimation of gene constraint from an evolutionary model with gene features”.
641 *Nature Genetics* (July 2024), pp. 1–12. doi: 10.1038/s41588-024-01820-9.
- 642 [62] David R. Kelley et al. “Sequential regulatory activity prediction across chromosomes with convolutional neu-
643 ral networks”. *Genome Research* 28.5 (Mar. 2018), pp. 739–750. doi: 10.1101/gr.227819.117.
- 644 [63] Vikram Agarwal and Jay Shendure. “Predicting mRNA Abundance Directly from Genomic Sequence Using
645 Deep Convolutional Neural Networks”. *Cell Reports* 31.7 (May 2020). doi: 10.1016/j.celrep.2020.107663.
- 646 [64] Žiga Avsec et al. “Effective gene expression prediction from sequence by integrating long-range interactions”.
647 *Nature Methods* 18.10 (Oct. 2021), pp. 1196–1203. doi: 10.1038/s41592-021-01252-x.
- 64 [65] Shiron Drusinsky, Sean Whalen, and Katherine S. Pollard. *Deep-learning prediction of gene expression from*
649 *personal genomes*. July 2024. doi: 10.1101/2024.07.27.605449.
- 650 [66] Atray Dixit et al. “Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled
651 Genetic Screens”. *Cell* 167.7 (Dec. 2016), 1853–1866.e17. doi: 10.1016/j.cell.2016.11.038.
- 652 [67] Ryan Tewhey et al. “Direct Identification of Hundreds of Expression-Modulating Variants using a Multiplexed
653 Reporter Assay”. *Cell* 165.6 (June 2016), pp. 1519–1529. doi: 10.1016/j.cell.2016.04.027.
- 654 [68] Martin Kircher et al. “Saturation mutagenesis of twenty disease-associated regulatory elements at single
655 base-pair resolution”. *Nature Communications* 10.1 (Aug. 2019), p. 3583. doi: 10.1038/s41467-019-11526-w.
- 656 [69] Jun Cheng et al. “Accurate proteome-wide missense variant effect prediction with AlphaMissense”. *Science*
657 381.6664 (Sept. 2023), eadg7492. doi: 10.1126/science.adg7492.
- 65 [70] Douglas M. Fowler and Stanley Fields. “Deep mutational scanning: a new style of protein science”. *Nature*
659 *Methods* 11.8 (July 2014), pp. 801–807. doi: 10.1038/nmeth.3027.
- 660 [71] GTEx Consortium. “Genetic effects on gene expression across human tissues”. *Nature* 550.7675 (Oct. 2017),
661 pp. 204–213. doi: 10.1038/nature24277.
- 662 [72] Min Zhao et al. “Computational tools for copy number variation (CNV) detection using next-generation se-
663 quencing data: features and perspectives”. *BMC Bioinformatics* 14.11 (Sept. 2013), S1. doi: 10.1186/1471-
664 2105-14-S11-S1.
- 665 [73] Xiang Zhu and Matthew Stephens. “Bayesian large-scale multiple regression with summary statistics from
666 genome-wide association studies”. *The Annals of Applied Statistics* 11.3 (Sept. 2017), pp. 1561–1592. doi: 10.
667 1214/17-AOAS1046.
- 66 [74] Heli Julkunen et al. “Atlas of plasma NMR biomarkers for health and disease in 118,461 individuals from the
669 UK Biobank”. *Nature Communications* 14.1 (Feb. 2023), p. 604. doi: 10.1038/s41467-023-36231-7.
- 670 [75] Clare Bycroft et al. “The UK Biobank resource with deep phenotyping and genomic data”. *Nature* 562.7726
671 (Oct. 2018), pp. 203–209. doi: 10.1038/s41586-018-0579-z.
- 672 [76] Gad Abraham, Yixuan Qiu, and Michael Inouye. “FlashPCA2: principal component analysis of Biobank-scale
673 genotype datasets”. *Bioinformatics* 33.17 (Sept. 2017), pp. 2776–2778. doi: 10.1093/bioinformatics/btx299.
- 674 [77] William McLaren et al. “The Ensembl Variant Effect Predictor”. *Genome Biology* 17.1 (June 2016), p. 122. doi:
675 10.1186/s13059-016-0974-4.
- 676 [78] Konrad J. Karczewski et al. “The mutational constraint spectrum quantified from variation in 141,456 hu-
677 mans”. *Nature* 581.7809 (May 2020), pp. 434–443. doi: 10.1038/s41586-020-2308-7.

- 67 [79] Sairam Behera et al. “Comprehensive genome analysis and variant detection at scale using DRAGEN”. *Nature*
679 *Biotechnology* (Oct. 2024), pp. 1–15. doi: 10.1038/s41587-024-02382-1.
- 6 0 [80] Adam Frankish et al. “GENCODE: reference annotation for the human and mouse genomes in 2023”. *Nucleic*
6 1 *Acids Research* 51.D1 (Jan. 2023), pp. D942–D949. doi: 10.1093/nar/gkac1071.
- 6 2 [81] Bernard Delyon, Marc Lavielle, and Eric Moulines. “Convergence of a Stochastic Approximation Version of
6 3 the EM Algorithm”. *The Annals of Statistics* 27.1 (Feb. 1999), pp. 94–128.
- 6 4 [82] Estelle Kuhn and Marc Lavielle. “Coupling a stochastic approximation version of EM with an MCMC proce-
6 5 dure”. *ESAIM: Probability and Statistics* 8 (Aug. 2004), pp. 115–131. doi: 10.1051/ps:2004007.
- 6 6 [83] Jeffrey P. Spence et al. *A flexible modeling and inference framework for estimating variant effect sizes from*
6 7 *GWAS summary statistics*. Apr. 2022. doi: 10.1101/2022.04.18.488696.
- 6 [84] Fabio Morgante et al. “A flexible empirical Bayes approach to multivariate multiple regression, and its im-
6 9 proved accuracy in predicting multi-tissue gene expression from genotypes”. *PLOS Genetics* 19.7 (July 2023),
690 e1010539. doi: 10.1371/journal.pgen.1010539.
- 691 [85] Deborah Kunkel et al. *Improving polygenic prediction from summary data by learning patterns of effect sharing*
692 *across multiple phenotypes*. May 2024. doi: 10.1101/2024.05.06.592745.
- 693 [86] Matthew Stephens. “False discovery rates: a new deal”. *Biostatistics* (Oct. 2016), kxw041. doi: 10.1093/
694 biostatistics/kxw041.

Appendices

1 A GENE DOSAGE RESPONSE CURVES

2 GENOME-WIDE BURDEN

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

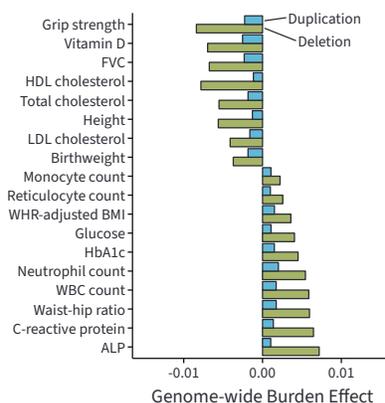


FIGURE A.1. Genome-wide burden data, reprocessed from work by Auwerx et al. [1]. These effects have a p-value of less than 0.05. The deletion or duplication effect is the effect of increasing numbers of deleted/duplicated genes on the trait.

7 SCALE OF THE DOSAGE AND TRAIT

9 The scale and shape of the curve depend on how gene expression and the trait are measured. We assume
10 that expression is measured on a logarithmic scale relative to the population mean. This allows us to
11 define the expression effect of loss-of-function (LoF) and duplication variants on a logarithmic scale such
12 that expression-decreasing variants have negative expression values and expression-increasing variants
13 have positive expression values. A heterozygous LoF genotype thus has an expression effect of $\log_2 \left(\frac{1}{2} \right) =$
14 -1 , and a duplication resulting in a copy number of three has an expression effect of $\log_2 \left(\frac{3}{2} \right) \approx 0.59$.
15 For traits, we chose to use an inverse rank normal transform. By centering the trait, we implicitly define
16 trait-decreasing variants as having negative effects and trait-increasing variants as having positive effects.

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

1 COMPOSITION OF MONOTONE CURVES

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

23 Now, suppose that g is a monotone increasing function. That is, $x \leq y$ implies that $g(x) \leq g(y)$
24 for any $x, y \in \mathbb{R}$. Similarly, suppose that f is a monotone increasing function. Then, $x \leq y$ implies that
25 $g(x) \leq g(y)$, which implies that $f(g(x)) \leq f(g(y))$. Thus, the composition $(f \circ g)$ is also monotone
26 increasing. Similar arguments for monotone decreasing functions or pairs of increasing and decreasing
27 functions demonstrate that the composition of monotone functions is also monotone.

2 B MONOTONICITY MODEL

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

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

37 We use the following linear system to model the phenotype:

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

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

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

41 DISTRIBUTION OF EFFECT SIZES

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

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

47 Suppose we collect the block matrices above into $\boldsymbol{\beta} \in \mathbb{R}^{2M}$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{2M \times 2M}$. Then, we can suc-
4 cinctly state that

$$\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

49 Let $\boldsymbol{\gamma}_j \in \mathbb{R}^2$ represent the latent effect sizes for the j th gene. For instance, the first coordinate may repre-
50 sent the LoF effect, and the second coordinate may represent the duplication effect. Then, the marginal
51 distribution is written as

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

52 PARAMETER OF INTEREST

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

$$\phi \triangleq -[\gamma_{1j}\gamma_{2j}].$$

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

$$|[\gamma_{1j}\gamma_{2j}]|^2 \leq [\gamma_{1j}^2][\gamma_{2j}^2],$$

58 so that

$$\phi \triangleq \frac{[\gamma_{1j}\gamma_{2j}]}{\sqrt{[\gamma_{1j}^2][\gamma_{2j}^2]}} = \frac{\bar{\gamma}_1\bar{\gamma}_2 + \sigma_{12}}{\sqrt{(\bar{\gamma}_1^2 + \sigma_{11}^2)(\bar{\gamma}_2^2 + \sigma_{22}^2)}}.$$

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

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

61 DISTRIBUTION OF GENOTYPES

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

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

$$\begin{aligned} \mathbf{X}_{ij} + 2p_j &\stackrel{\text{iid}}{\sim} \text{Binomial}(2, p_j) \\ \mathbf{Z}_{ij} + 2q_j &\stackrel{\text{iid}}{\sim} \text{Binomial}(2, q_j). \end{aligned}$$

72 There is no LD between variant classes,

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

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

$$\begin{aligned} [\mathbf{X}^\top \mathbf{X}] &= \mathbf{P}^{\frac{1}{2}} \mathbf{R}_1 \mathbf{P}^{\frac{1}{2}} \\ [\mathbf{Z}^\top \mathbf{Z}] &= \mathbf{Q}^{\frac{1}{2}} \mathbf{R}_2 \mathbf{Q}^{\frac{1}{2}}, \end{aligned}$$

76 where

$$\begin{aligned} \mathbf{P} &= \text{diag}(\{P_j\}_{j=1}^M) \\ \mathbf{Q} &= \text{diag}(\{Q_j\}_{j=1}^M). \end{aligned}$$

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

0 BURDEN REGRESSION

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

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

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

$$\begin{aligned} \hat{\gamma}_{1j} &= (\mathbf{X}_{\cdot j}^\top \mathbf{X}_{\cdot j})^{-1} \mathbf{X}_{\cdot j}^\top \mathbf{Y} \approx \frac{1}{NP_j} \mathbf{X}_{\cdot j}^\top \mathbf{Y} \\ \hat{\gamma}_{2j} &= (\mathbf{Z}_{\cdot j}^\top \mathbf{Z}_{\cdot j})^{-1} \mathbf{Z}_{\cdot j}^\top \mathbf{Y} \approx \frac{1}{NQ_j} \mathbf{Z}_{\cdot j}^\top \mathbf{Y}. \end{aligned}$$

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

$$\begin{aligned} \hat{s}_{1j} &\triangleq \widehat{\text{SE}}(\hat{\gamma}_{1j}) = \sqrt{\sigma_{\epsilon_1}^2 (\mathbf{X}_{\cdot j}^\top \mathbf{X}_{\cdot j})^{-1}} \approx \frac{1}{\sqrt{NP_j}} \\ \hat{s}_{2j} &\triangleq \widehat{\text{SE}}(\hat{\gamma}_{2j}) = \sqrt{\sigma_{\epsilon_2}^2 (\mathbf{Z}_{\cdot j}^\top \mathbf{Z}_{\cdot j})^{-1}} \approx \frac{1}{\sqrt{NQ_j}}. \end{aligned}$$

7 Therefore, the Z scores are

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

C MONOTONICITY INFERENCE

9 MAXIMUM LIKELIHOOD ESTIMATION

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

$$\sim \mathcal{N}(\boldsymbol{\theta}, \boldsymbol{\Sigma}(\boldsymbol{\omega})),$$

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

94 Next, the likelihood of the estimated effect sizes is defined using the approximate regression with
 95 summary statistics (RSS) likelihood [3]. Under this likelihood, the estimated effect sizes are patterned
 96 by the LD in the cohort. Furthermore, the standard error of the sampling distribution is used as an es-
 97 timate of the dispersion around the mean. Zhu et al. showed that this likelihood asymptotically ap-
 9 approaches the sampling distribution of the estimated effect sizes [3]. Let $\hat{\mathbf{S}}_1 = \text{diag}(\{s_{1j}\}_{j=1}^M)$ and
 99 $\hat{\mathbf{S}}_2 = \text{diag}(\{s_{2j}\}_{j=1}^M)$, and let $\hat{\mathbf{R}}_1$ and $\hat{\mathbf{R}}_2$ represent the in-sample correlation matrices of the two variant
 100 classes respectively. Then, in notation reflecting the RSS likelihood, we define

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

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

101 Then, the RSS likelihood for the observed effect sizes is

$$\hat{\boldsymbol{\alpha}} \mid \sim \mathcal{N}(\hat{\mathbf{S}}\hat{\mathbf{R}}\hat{\boldsymbol{\alpha}}, \hat{\mathbf{S}}\hat{\mathbf{R}}\hat{\boldsymbol{\alpha}}).$$

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

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

107 In practice, these are derived by dropping small eigenvalues from the numerical eigendecomposition of
 10 $\hat{\mathbf{S}}\hat{\mathbf{R}}\hat{\mathbf{S}}$. Rather than modeling the observations $\hat{\boldsymbol{\alpha}}$, we model the projected data,

$$\mathbf{U}^\top \hat{\boldsymbol{\alpha}} \mid \sim \mathcal{N}(\mathbf{U}^\top \hat{\mathbf{S}}\hat{\mathbf{R}}\hat{\boldsymbol{\alpha}}, \mathbf{U}^\top \hat{\mathbf{S}}\hat{\mathbf{R}}\hat{\boldsymbol{\alpha}}).$$

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

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

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

$$\mathbf{U}^{\top \wedge} \sim \mathcal{N}(\mathbf{A}^{-}, \mathbf{K}).$$

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

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

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

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

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

116 INFERENCE

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

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

119 The log likelihood is

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

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

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

121 For maximum likelihood estimation of θ and ω , we use natural gradient ascent [4]. This is an opti-
 122 mization approach that uses both first- and second-order information about the log likelihood via the
 123 gradient and Fisher information matrix. The gradient of the log likelihood function with respect to θ is

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

124 Using the matrix chain rule, the derivative of the log likelihood with respect to one of the mean paramete-
 125 rs is

$$\frac{\partial \ell}{\partial \theta_i} = \text{Tr} \left[(\nabla_{\theta} \ell)^{\top} \frac{\partial \theta}{\partial \theta_i} \right] = (\nabla_{\theta} \ell)^{\top} \frac{\partial \theta}{\partial \theta_i}.$$

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

$$\nabla_{\theta} \ell = (\nabla_{\theta} \ell)^{\top} \frac{\partial \theta}{\partial \theta}.$$

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

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

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

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

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

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

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

$$(\theta, \omega) = \begin{bmatrix} -(\theta, \omega) & \mathbf{0} \\ \mathbf{0} & \mathbf{K}(\omega) \end{bmatrix}.$$

132 The Fisher information of the mean parameter is

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

133 The Fisher information of the covariance parameters is

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

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

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{t+1} &= \hat{\boldsymbol{\theta}}_t + \alpha_t \left[-(\hat{\boldsymbol{\theta}}_t, \hat{\boldsymbol{\omega}}_t) \right]^{-1} \left[\nabla_{\boldsymbol{\theta}} \ell \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_t} \right] \\ \hat{\boldsymbol{\omega}}_{t+1} &= \hat{\boldsymbol{\omega}}_t + \alpha_t \left[\mathbf{K}(\hat{\boldsymbol{\omega}}_t) \right]^{-1} \left[\nabla_{\boldsymbol{\omega}} \ell \Big|_{\boldsymbol{\omega}=\hat{\boldsymbol{\omega}}_t} \right]. \end{aligned}$$

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

140 UNCERTAINTY ESTIMATION

141 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
 142 continuous map from the parameters to a real number. ψ can itself be interpreted as a parameter of the
 143 model. In our model, ϕ is an example of such a parameter. A natural estimate for ψ is to map the MLEs,

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

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

$$\begin{aligned} \hat{\boldsymbol{\theta}} &\xrightarrow{d} \mathcal{N}(\boldsymbol{\theta}, [\mathbf{I} - \mathbf{K}(\boldsymbol{\omega})]^{-1}) \\ \hat{\boldsymbol{\omega}} &\xrightarrow{d} \mathcal{N}(\boldsymbol{\omega}, [\mathbf{K}(\boldsymbol{\omega})]^{-1}) \end{aligned}$$

145 By the delta method, the estimator for ψ converges to

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

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

14 FIXED-EFFECTS MODEL

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

151 that the observations have the following likelihood:

$$\mathbf{U}^\top \hat{\mathbf{y}} \mid \mathbf{A} \sim \mathcal{N}(\mathbf{A} \boldsymbol{\beta}, \boldsymbol{\Sigma}).$$

152 For unbiased estimation, the projection matrix $\mathbf{A}^+ \mathbf{A}$ is used frequently throughout our derivations.

153 To see that it is a projection matrix, note that

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

154 Then, it follows that

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

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

156 MEAN EFFECT

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

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

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

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

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

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

$$\begin{aligned} \mathbb{E}\left[\begin{bmatrix} \mathbf{1} \\ \mathbf{0} \end{bmatrix}^\top \mathbf{A}^+ \mathbf{U}^\top \hat{\gamma}\right] &= \begin{bmatrix} \mathbf{1} \\ \mathbf{0} \end{bmatrix}^\top \mathbf{A}^+ \mathbb{E}\left[\mathbf{U}^\top \hat{\gamma}\right] \\ &= \begin{bmatrix} \mathbf{1} \\ \mathbf{0} \end{bmatrix}^\top \mathbf{A}^+ \mathbf{A} \\ &= \begin{bmatrix} \mathbf{1} \\ \mathbf{0} \end{bmatrix}^\top \mathbf{A}^+ \mathbf{A} \left(\begin{bmatrix} \mathbf{1} \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{2} \end{bmatrix} \right) \\ &= \begin{bmatrix} \mathbf{1} \\ \mathbf{0} \end{bmatrix}^\top \mathbf{A}^+ \mathbf{A} \begin{bmatrix} \mathbf{1} \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{1} \\ \mathbf{0} \end{bmatrix}^\top \mathbf{A}^+ \mathbf{A} \begin{bmatrix} \mathbf{0} \\ \mathbf{2} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{1} \\ \mathbf{0} \end{bmatrix}^\top \mathbf{A}^+ \mathbf{A} \begin{bmatrix} \mathbf{1} \\ \mathbf{0} \end{bmatrix}. \end{aligned}$$

165 The last equality holds because $\mathbf{A}^+ \mathbf{A}$ is block diagonal. Note here that $\mathbf{A}^+ \mathbf{A}$ is a projection matrix and
 166 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

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

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

170 MEAN SQUARED EFFECT

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

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

174 Since the second moment of a univariate normal random variable is

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

175 we note that

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

176 is an unbiased estimate of the squared burden effect size.

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

1 0 **RANDOM-EFFECTS MODEL**

1 1 A random-effects model is required to develop estimators for the covariance components. We use an
 1 2 approach inspired by linkage disequilibrium score regression (LDSC) [9], and similar to that outlined in
 1 3 [3]. We assume that \mathbf{z} is random, with finite first and second moments. We also assume that the latent
 1 4 effect sizes are independent given $\hat{\mathbf{S}}$ and $\hat{\mathbf{R}}$. We make the following minimal assumptions about first and
 1 5 second moments of the joint distribution of the latent effect sizes:

$$\begin{aligned} \mathbb{E}[\mathbf{z} \mid \hat{\mathbf{S}}, \hat{\mathbf{R}}] &= \mathbf{0} \\ \mathbb{V}[\mathbf{z} \mid \hat{\mathbf{S}}, \hat{\mathbf{R}}] &= \mathbf{\Sigma}. \end{aligned}$$

1 6 The Z scores from the burden regression can be defined as

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

1 7 By the properties of the normal distribution, the Z scores are distributed as

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

1 and Z scores for the two variant classes are independent,

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

1 9 Without loss of generality for the second variant class, we focus on the j th gene of the first variant class.
 190 Let \hat{r}_{1ji} be the i th element of $\hat{\mathbf{R}}_1$. The marginal distribution of the Z score for a given gene is then

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

191 We define the scaled LD score for the j th gene as

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

192 It is also useful to define

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

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

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

194 We define the cross-variant LD score as

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

195 It is also useful to define

$$l_j \triangleq \begin{bmatrix} \sum_{i=1}^M \frac{\hat{r}_{1ji}}{\hat{s}_{1j}} \\ \sum_{i=1}^M \frac{\hat{r}_{2ji}}{\hat{s}_{2j}} \end{bmatrix}.$$

196 COVARIANCE COMPONENTS

197 Similar to LDSC, we now proceed by evaluating the expected squared Z score to develop estimators for
 19 σ_{11}^2 and σ_{22}^2 . We integrate over the latent effect size, γ , using the law of total expectation and the law of
 199 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
 200 expected squared Z score is

$$\begin{aligned} [z_{1j}^2] &= \mathbb{V}[z_{1j}] + ([z_{1j}])^2 \\ &= \mathbb{V}[z_{1j} \mid \cdot] + [\mathbb{V}[z_{1j} \mid \cdot]] + ([z_{1j} \mid \cdot])^2. \end{aligned}$$

201 The first term introduces the LD score into the expectation,

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

202 The second term introduces a constant,

$$[\mathbb{V}[z_{1j} | \cdot]] = 1.$$

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

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

204 Taken together, the expected squared Z score is

$$[z_{1j}^2] = \sigma_{11}^2 l_{1j} + \bar{\gamma}_1^2 l_{1j}^2 + 1.$$

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

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

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

$$\begin{aligned} [z_{1j} z_{2j}] &= \mathbb{V}[z_{1j}, z_{2j}] + [z_{1j}] [z_{2j}] \\ &= \mathbb{V}([z_{1j} | \cdot], [z_{2j} | \cdot]) + [\mathbb{V}[z_{1j}, z_{2j} | \cdot]] + [[z_{1j} | \cdot]] [[z_{2j} | \cdot]]. \end{aligned}$$

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

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

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

$$[\mathbb{V} [z_{1j}, z_{2j} |]] = 0.$$

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

$$\begin{aligned} [[z_{1j} |]] [[z_{2j} |]] &= \begin{bmatrix} \sum_{i=1}^M \frac{\hat{r}_{1ji}}{\hat{s}_{1i}} \gamma_{1i} \end{bmatrix} \begin{bmatrix} \sum_{k=1}^M \frac{\hat{r}_{2jk}}{\hat{s}_{2k}} \gamma_{2k} \end{bmatrix} \\ &= \begin{bmatrix} \sum_{i=1}^M \frac{\hat{r}_{1ji}}{\hat{s}_{1i}} [\gamma_{1i}] \end{bmatrix} \begin{bmatrix} \sum_{k=1}^M \frac{\hat{r}_{2jk}}{\hat{s}_{2k}} [\gamma_{2k}] \end{bmatrix} \\ &= \bar{\gamma}_1 \bar{\gamma}_2 l_{1j} l_{2j}. \end{aligned}$$

216 Taken together, the expected product of Z scores is

$$[z_{1j} z_{2j}] = \sigma_{12} \ell_j + \bar{\gamma}_1 \bar{\gamma}_2 l_{1j} l_{2j}.$$

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

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

21 MONOTONICITY

219 These estimators of individual parameters in the model allow us to define the MoM estimator for mono-
220 tonicity as

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

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

224 D SIMULATIONS

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

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

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

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

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

246 UNBIASED AVERAGE BURDEN EFFECT ESTIMATORS

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

It is made available under a [CC-BY-NC 4.0 International license](https://creativecommons.org/licenses/by-nc/4.0/).

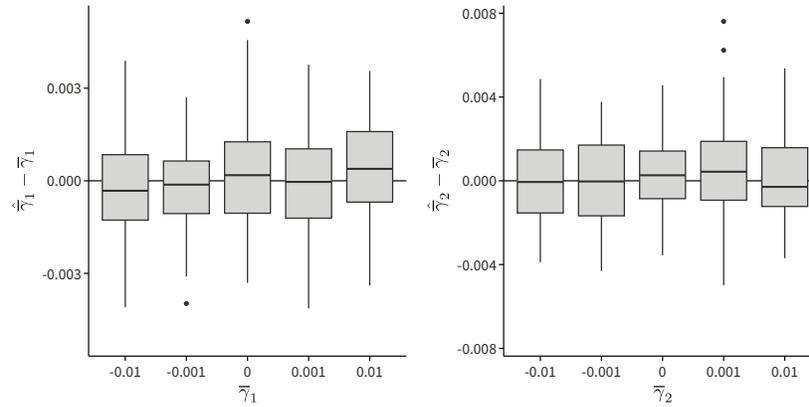


FIGURE D.1. (Left) The difference in the estimated value $\hat{\gamma}_1$ and the true value $\bar{\gamma}_1$ for various tested values are shown for 100 simulations. (Right) The difference in the estimated value $\hat{\gamma}_2$ and the true value $\bar{\gamma}_2$ for various tested values are shown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

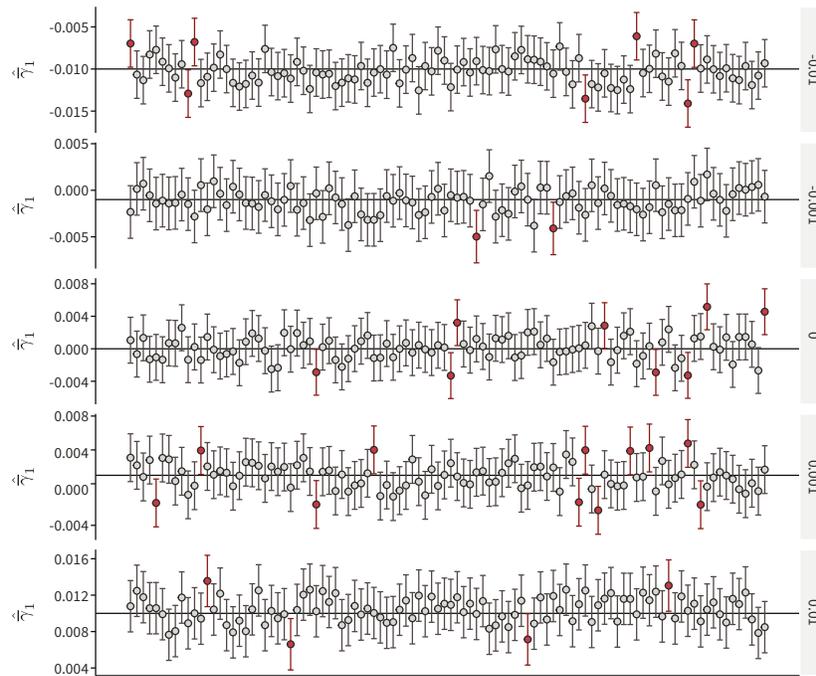


FIGURE D.2. The plot shows 95% confidence intervals from 100 simulations for various values of $\bar{\gamma}_1$. Confidence intervals in red represent realizations that do not cover the true value.

It is made available under a [CC-BY-NC 4.0 International license](https://creativecommons.org/licenses/by-nc/4.0/).

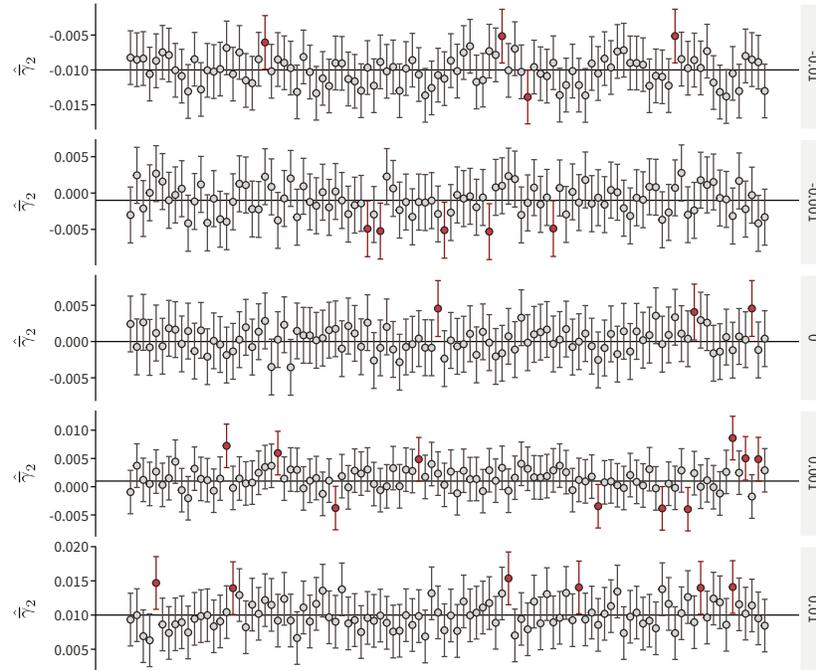


FIGURE D.3. The plot shows 95% confidence intervals from 100 simulations for various values of $\bar{\gamma}_2$. Confidence intervals in red represent realizations that do not cover the true value.

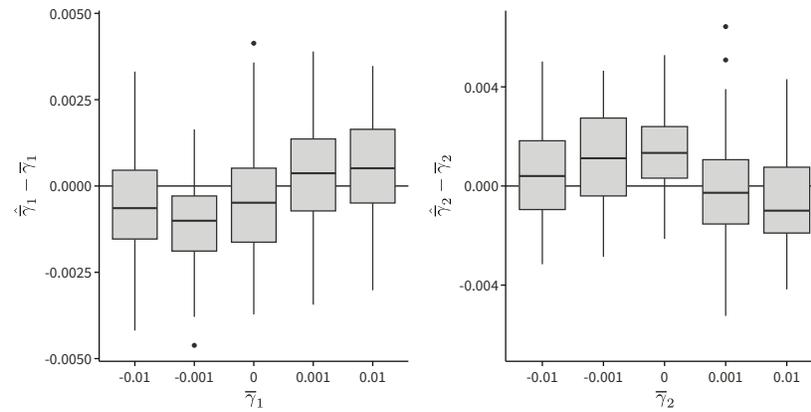


FIGURE D.4. (Left) The difference in the estimated value $\hat{\gamma}_1$ and the true value $\bar{\gamma}_1$ for various tested values are shown for 100 simulations. (Right) The difference in the estimated value $\hat{\gamma}_2$ and the true value $\bar{\gamma}_2$ for various tested values are shown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

It is made available under a [CC-BY-NC 4.0 International license](https://creativecommons.org/licenses/by-nc/4.0/).

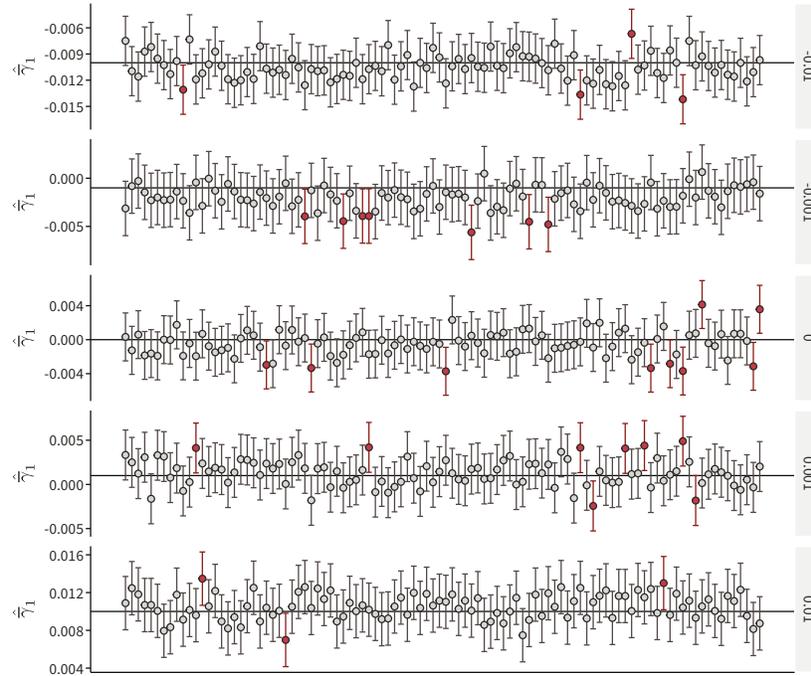


FIGURE D.5. The plot shows 95% confidence intervals from 100 simulations for various values of γ_1 . Confidence intervals in red represent realizations that do not cover the true value.

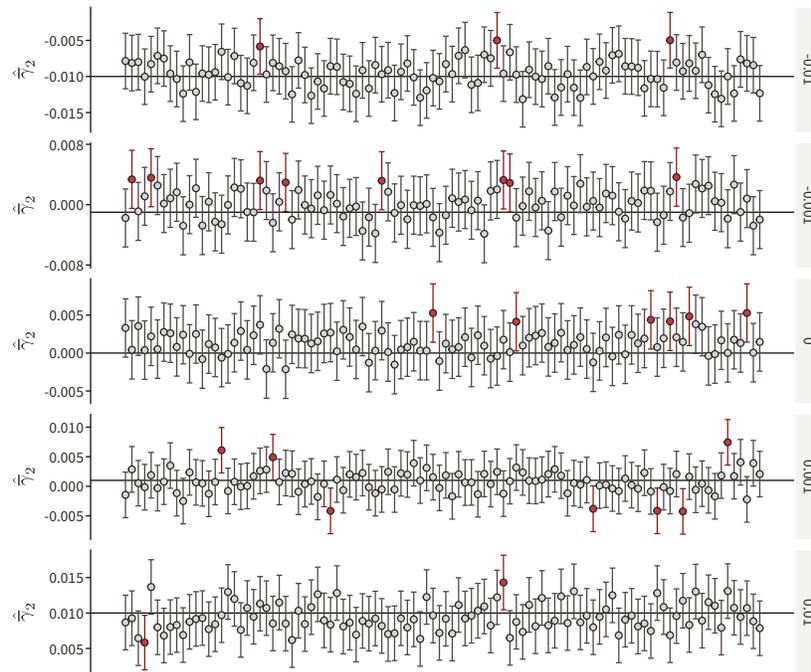


FIGURE D.6. The plot shows 95% confidence intervals from 100 simulations for various values of γ_2 . Confidence intervals in red represent realizations that do not cover the true value.

253 UNBIASED AVERAGE SQUARED BURDEN EFFECT ESTIMATOR

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

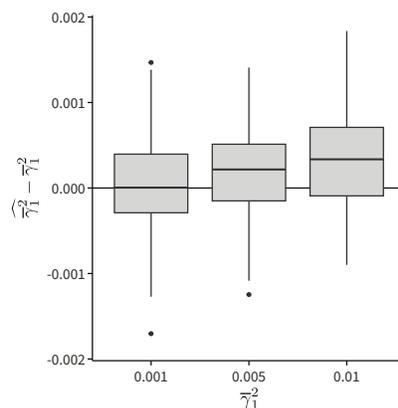


FIGURE D.7. The difference in the estimated value $\widehat{\bar{\gamma}}_1^2$ and the true value $\bar{\gamma}_1^2$ for various tested values are shown for 100 simulations.

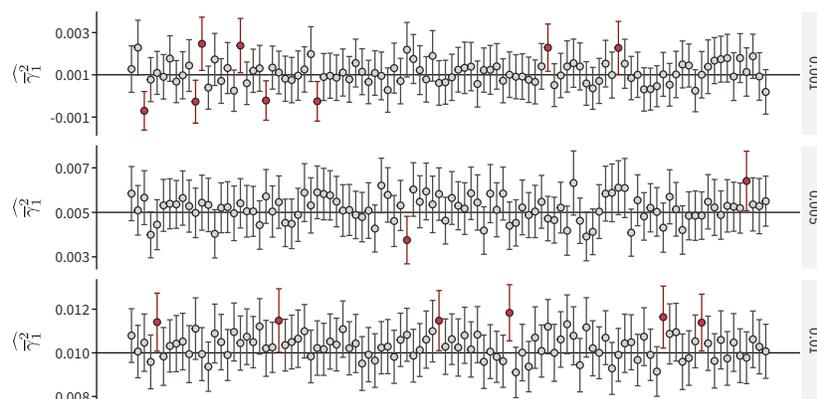


FIGURE D.8. The plot shows 95% confidence intervals from 100 simulations for various values of $\bar{\gamma}_1^2$. Confidence intervals in red represent realizations that do not cover the true value.

257 METHOD-OF-MOMENTS MONOTONICITY ESTIMATOR

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

It is made available under a [CC-BY-NC 4.0 International license](https://creativecommons.org/licenses/by-nc/4.0/) .

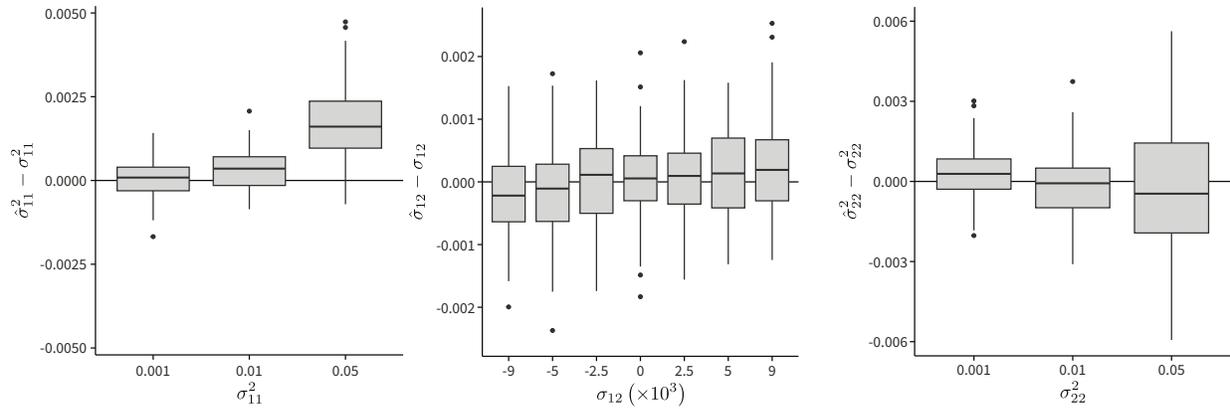


FIGURE D.9. (Left) The difference in the estimated value $\hat{\sigma}_{11}^2$ and the true value σ_{11}^2 for various tested values are shown for 100 simulations. (Center) The difference in the estimated value $\hat{\sigma}_{12}$ and the true value σ_{12} for various tested values are shown for 100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

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

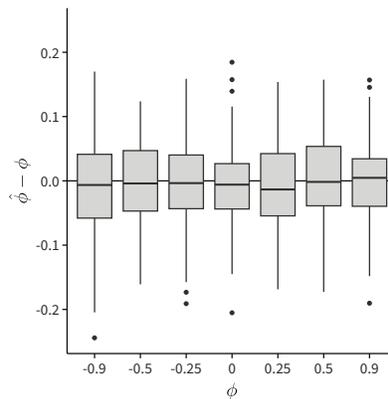


FIGURE D.10. The difference in the estimated value $\hat{\phi}$ and the true value ϕ for various tested values are shown for 100 simulations.

It is made available under a [CC-BY-NC 4.0 International license](https://creativecommons.org/licenses/by-nc/4.0/).

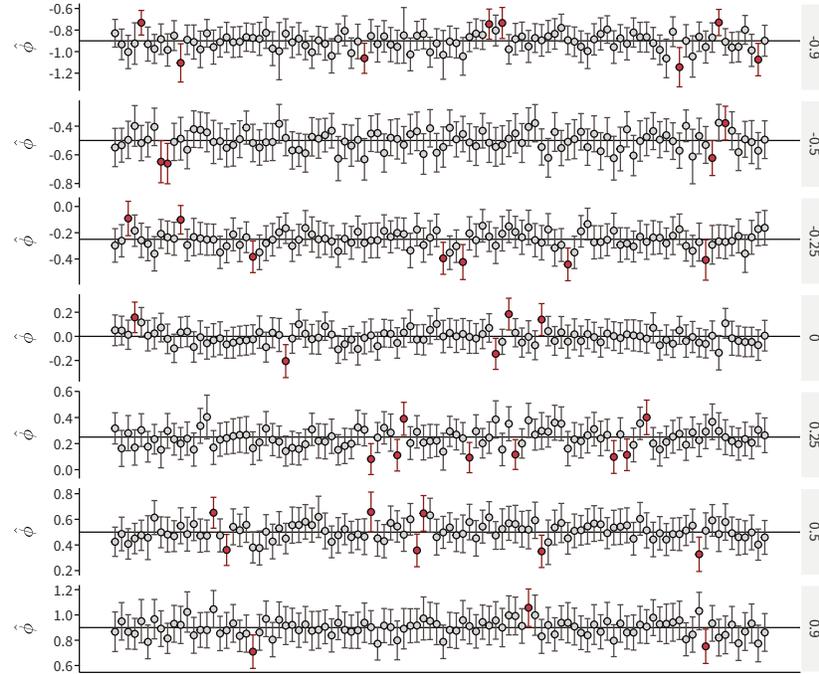


FIGURE D.11. The plot shows 95% confidence intervals from 100 simulations for various values of ϕ . Confidence intervals in red represent realizations that do not cover the true value.

265 **MAXIMUM LIKELIHOOD AVERAGE BURDEN EFFECT ESTIMATORS**

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

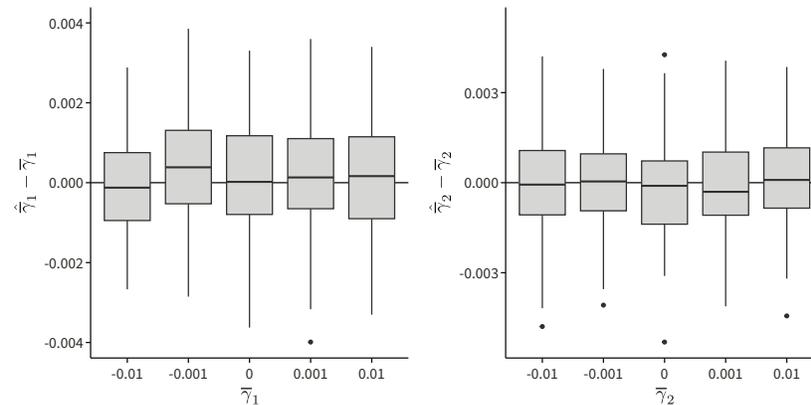


FIGURE D.12. (Left) The difference in the estimated value $\hat{\gamma}_1$ and the true value γ_1 for various tested values are shown for 100 simulations. (Right) The difference in the estimated value $\hat{\gamma}_2$ and the true value γ_2 for various tested values are shown for 100 simulations.

It is made available under a [CC-BY-NC 4.0 International license](https://creativecommons.org/licenses/by-nc/4.0/).

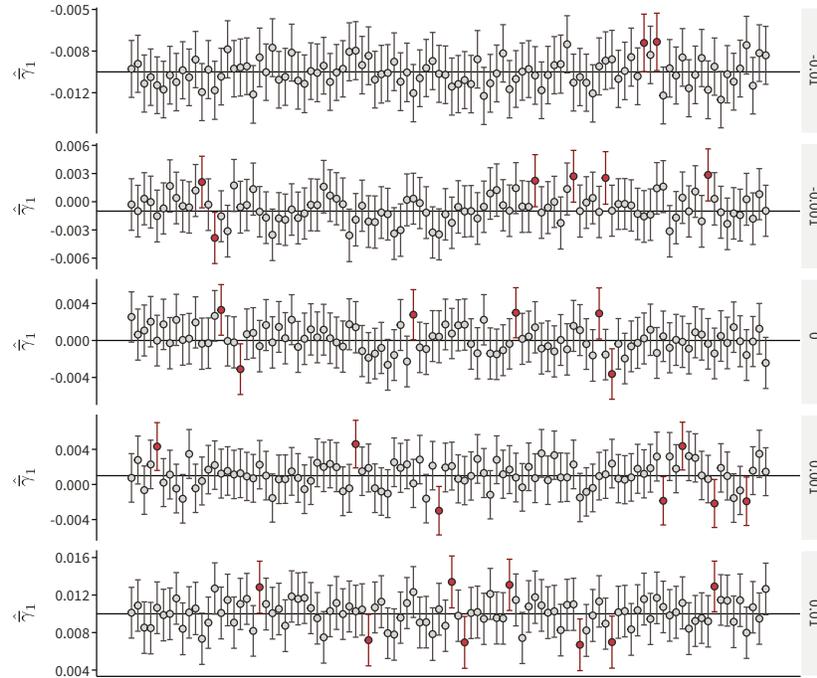


FIGURE D.13. The plot shows 95% confidence intervals from 100 simulations for various values of $\bar{\gamma}_1$. Confidence intervals in red represent realizations that do not cover the true value.

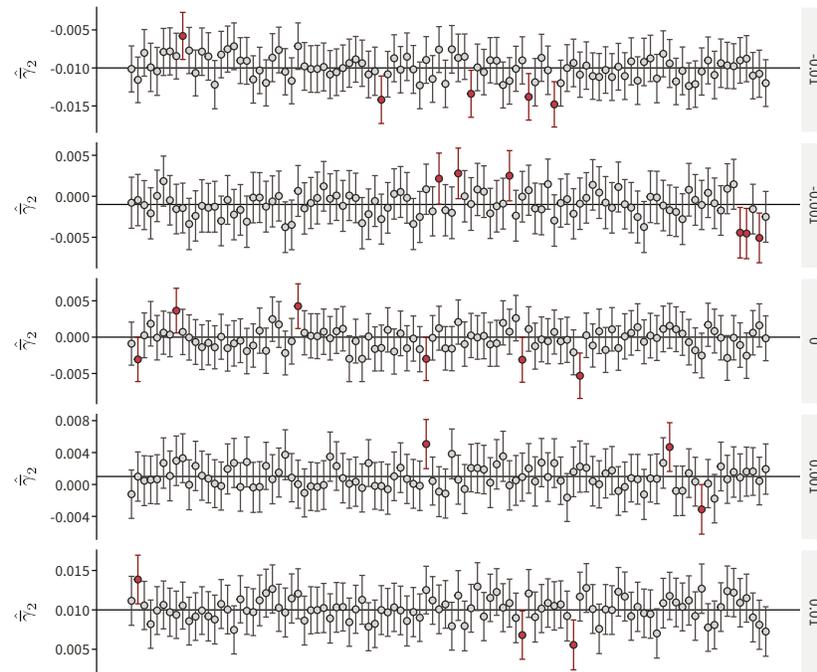


FIGURE D.14. The plot shows 95% confidence intervals from 100 simulations for various values of $\bar{\gamma}_2$. Confidence intervals in red represent realizations that do not cover the true value.

269 MAXIMUM LIKELIHOOD MONOTONICITY ESTIMATOR

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

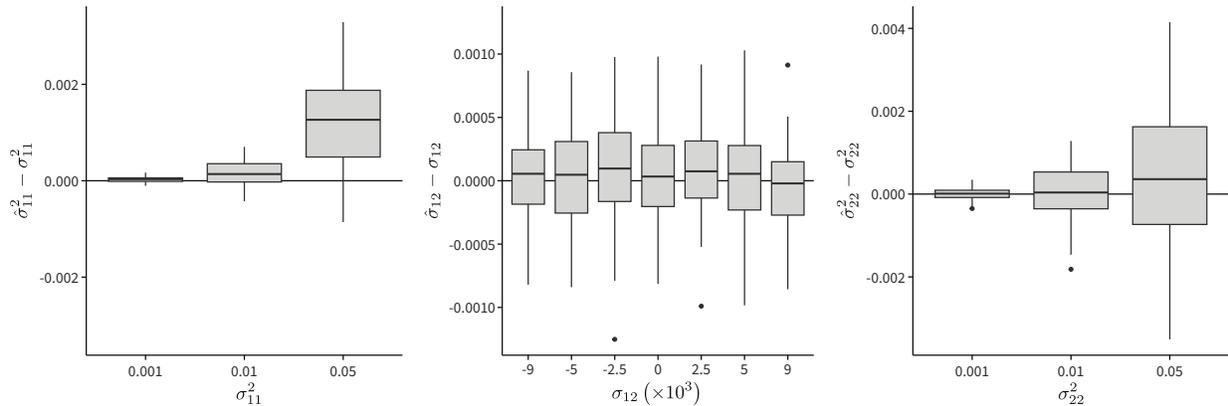


FIGURE D.15. (Left) The difference in the estimated value $\hat{\sigma}_{11}^2$ and the true value σ_{11}^2 for various tested values are shown for 100 simulations. (Center) The difference in the estimated value $\hat{\sigma}_{12}$ and the true value σ_{12} for various tested values are shown for 100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

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

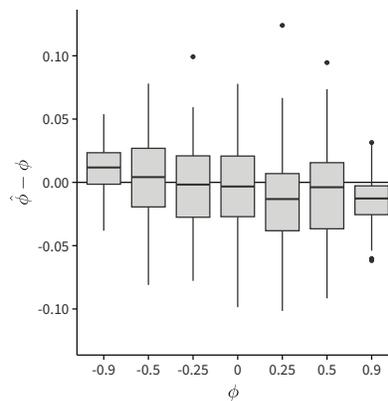


FIGURE D.16. The difference in the estimated value $\hat{\phi}$ and the true value ϕ for various tested values are shown for 100 simulations.

It is made available under a [CC-BY-NC 4.0 International license](https://creativecommons.org/licenses/by-nc/4.0/).

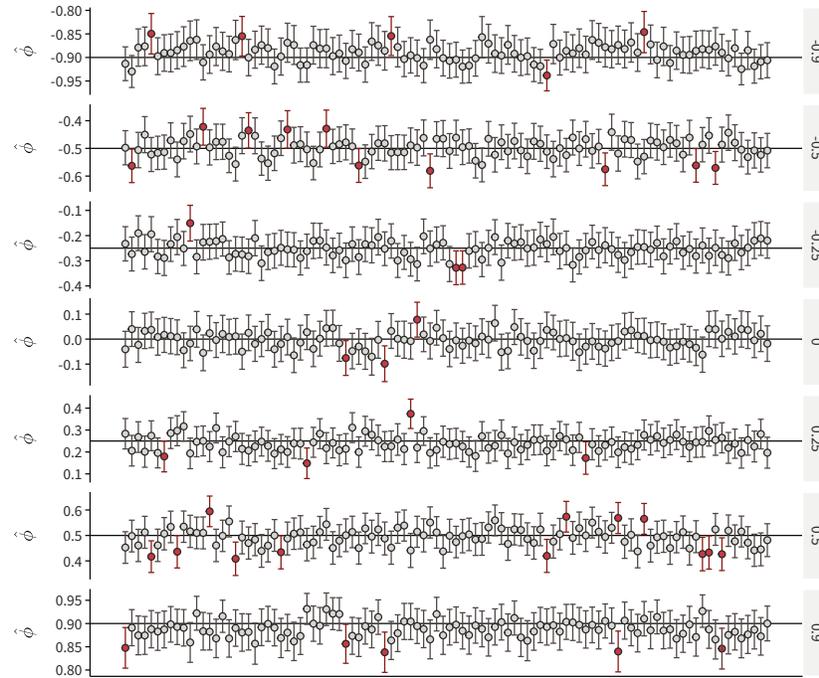


FIGURE D.17. The plot shows 95% confidence intervals from 100 simulations for various values of ϕ . Confidence intervals in red represent realizations that do not cover the true value.

27 E BUFFERING MODEL

279 In the buffering model, we are interested in evaluating if GDRCs are systematically buffered against one
 2 0 trait direction. This could occur if non-monotone GDRCs are preferentially present in one direction over
 2 1 the other. This can also occur if monotone GDRCs achieve larger values for the trait in one direction com-
 2 2 pared to the other.

2 3 We use the multivariate adaptive shrinkage (MASH) model [12] as a flexible prior for our latent effect
 2 4 sizes. The MASH prior consists of a mixture of multivariate normal distributions. Suppose $\{\mathbf{V}_k\}_{k=1}^K$ rep-
 2 5 represents a fixed set of covariance matrices in $\mathbb{R}^{2 \times 2}$ for K mixture components. We use bivariate covariance
 2 6 matrices over a grid of variances and correlation values. Let $\boldsymbol{\pi} \in \mathbb{R}^{K-1}$ represent the mixture weights.
 2 7 Then, the gene-level prior is

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

2 The prior over the entire set of M genes is

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

2 9 The likelihood model remains the same as before.

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

290 The joint likelihood is

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

291 INFERENCE USING EMPIRICAL BAYES

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

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

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

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

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

301 Since the analytical form of $Q(\pi | \pi^{(t)})$ is also intractable under our model, we use SAEM to approximate
 302 the expectation. We sample $\pi^{(1)}, \dots, \pi^{(B)} \stackrel{\text{iid}}{\sim} p(\cdot | \mathbf{U}^{\top \wedge}, \pi = \pi^{(t)})$ using Hamiltonian Monte Carlo [16,
 303 17]. Then, we approximate the expectation with

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

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

$$\tilde{Q}(\pi | \pi^{(t)}) \triangleq (1 - c_t) \tilde{Q}(\pi | \pi^{(t-1)}) + \frac{c_t}{B} \sum_{b=1}^B \log p(\mathbf{U}^{\top \wedge}, \pi^{(b)} | \pi = \pi^{(t)})$$

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

$$\pi^{(t+1)} = \arg \max_{\pi} \tilde{Q}(\pi | \pi^{(t)}).$$

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

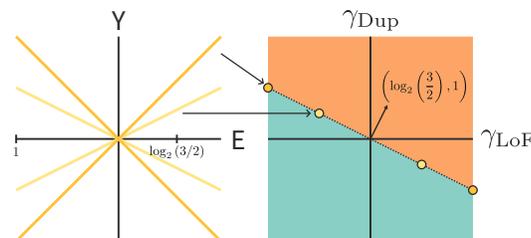


FIGURE E.1. The boundary between buffering curves is represented by the yellow gene dosage response curves on the left. These map to a diagonal across the burden effect plot. To estimate signal in one direction of the diagonal versus the other, we use the normal vector in black shown on the right.

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

315 $p \cdot | \mathbf{U}^{\top \wedge}, \boldsymbol{\pi} = \hat{\boldsymbol{\pi}})$. The estimand of interest is

$$\xi \triangleq \left[\log_2 \left(\frac{3}{2} \right) \gamma_{1j} + \gamma_{2j} \mid \mathbf{U}^{\top \wedge} \right] \approx \frac{1}{BM} \sum_{b=1}^B \sum_{j=1}^M \log_2 \left(\frac{3}{2} \right) \gamma_{1j}^{(b)} + \gamma_{2j}^{(b)},$$

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

$$\begin{aligned} \xi_{nm} \mathbb{P}(\gamma_{1j} \gamma_{2j} > 0) &= \left[\log_2 \left(\frac{3}{2} \right) \gamma_{1j} + \gamma_{2j} \mid \mathbf{U}^{\top \wedge}, \gamma_{1j} \gamma_{2j} > 0 \right] \mathbb{P}(\gamma_{1j} \gamma_{2j} > 0) \\ &\approx \frac{1}{BM} \sum_{b=1}^B \sum_{\gamma_{1j}^{(b)} \gamma_{2j}^{(b)} > 0} \log_2 \left(\frac{3}{2} \right) \gamma_{1j}^{(b)} + \gamma_{2j}^{(b)} \\ \xi_m \mathbb{P}(\gamma_{1j} \gamma_{2j} < 0) &= \left[\log_2 \left(\frac{3}{2} \right) \gamma_{1j} + \gamma_{2j} \mid \mathbf{U}^{\top \wedge}, \gamma_{1j} \gamma_{2j} < 0 \right] \mathbb{P}(\gamma_{1j} \gamma_{2j} < 0) \\ &\approx \frac{1}{BM} \sum_{b=1}^B \sum_{\gamma_{1j}^{(b)} \gamma_{2j}^{(b)} < 0} \log_2 \left(\frac{3}{2} \right) \gamma_{1j}^{(b)} + \gamma_{2j}^{(b)}. \end{aligned}$$

320 **F DATA ANALYSIS**

321 **SYNONYMOUS VARIANTS**

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

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

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

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

345 **NUMBER OF GENOTYPING PRINCIPAL COMPONENTS**

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

It is made available under a [CC-BY-NC 4.0 International license](https://creativecommons.org/licenses/by-nc/4.0/).

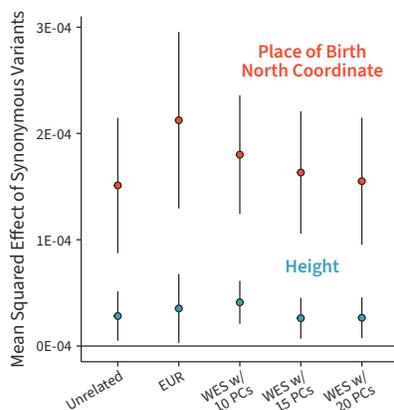


FIGURE F.1. The mean squared effect of synonymous variants for various cohorts of individuals and different numbers of genotyping PCs. The unrelated and EUR cohort burden tests used 15 genotyping PCs. We tested 10, 15, and 20 genotyping PCs in the WES cohort burden tests.

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

356 INFLATION FROM CONFOUNDING

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

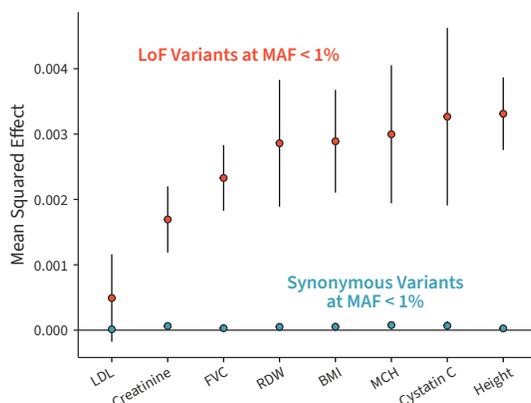


FIGURE F.2. LoF variants have a magnitude larger effect on traits than synonymous variants across various continuous traits.

364 **UTILITY OF MISANNOTATION PROBABILITY**

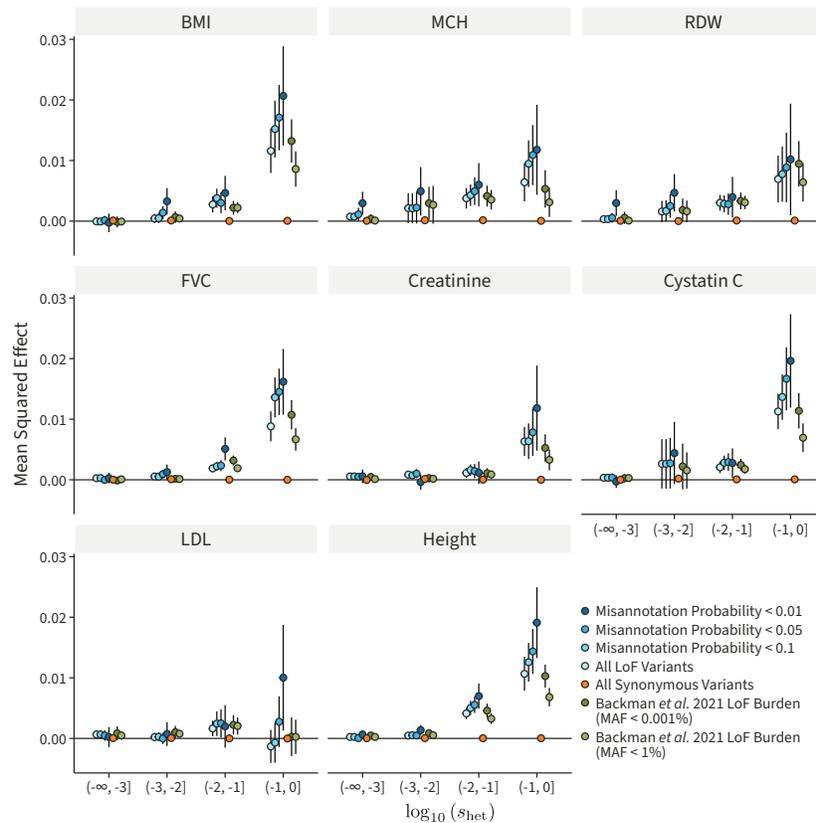


FIGURE F.3. The mean squared effect of genes in various buckets of $\log_{10}(s_{het})$ for various traits. Genes with larger effects tend to be more constrained. All data here uses a MAF < 1% filter other than the MAF < 0.001% associations from [19]. The use of the misannotation probability increases 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.

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

It is made available under a [CC-BY-NC 4.0 International license](https://creativecommons.org/licenses/by-nc/4.0/).

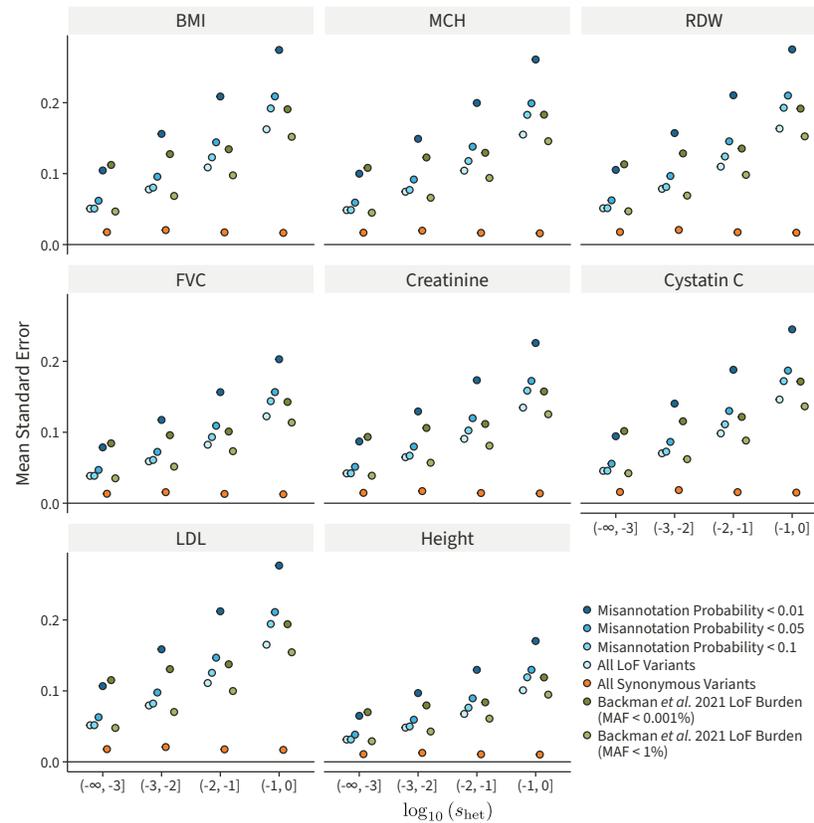


FIGURE F.4. The mean standard error of burden estimates for genes in various buckets of $\log_{10}(s_{het})$ for various traits. Since larger-effect variants are under increased selective constraint, their frequencies are lower and the estimation noise is larger. All data here uses a MAF < 1% filter 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.

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

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

3 4 **CNV GENOTYPING ERROR**

3 5 When we first estimated the LD matrices for the duplication burden genotypes, we noticed large amounts
3 6 of long-range LD (Figure F.5).

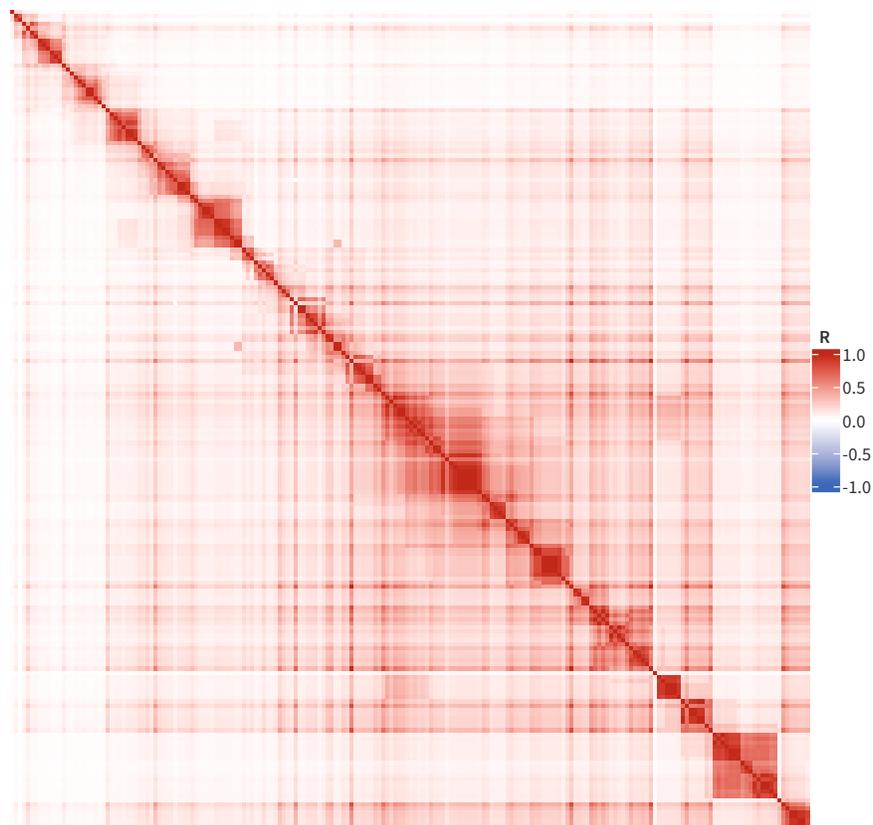


FIGURE F.5. The correlation matrix for the first 200 genes on chromosome 3 using all the samples. Although segregating duplications can be large, they are not expected to capture so many genes.

3 7 We reasoned that this is driven by a few poor-quality samples with systematically higher copy num-
3 ber. In their initial report on the UKB whole-genome sequencing (WGS) data, Li et al. reported an average
3 9 structural variant (SV) burden of 3.6 Mbp per haploid genome [22]. Assuming that the total copy number
390 variant (CNV) burden is on a similar order of magnitude, we decided to exclude any samples with more
391 than 10 Mbp of either deletion or duplication burden. Most samples had less than 10 Mbp of duplications
392 or deletions. However, some samples had upwards of 100 Mbp of affected sequence (Figure F.6).

It is made available under a [CC-BY-NC 4.0 International license](https://creativecommons.org/licenses/by-nc/4.0/) .

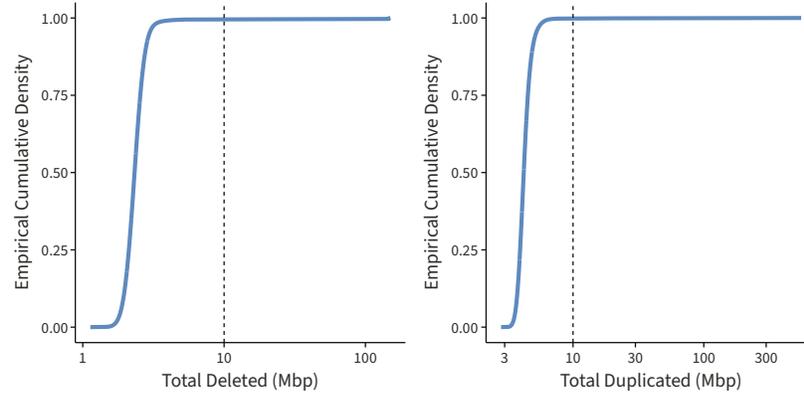


FIGURE F.6. The empirical cumulative density function (eCDF) of the total amount of genome sequence affected by either deletion or duplication. Our 10 Mbp filtering step impacted only a few samples.

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

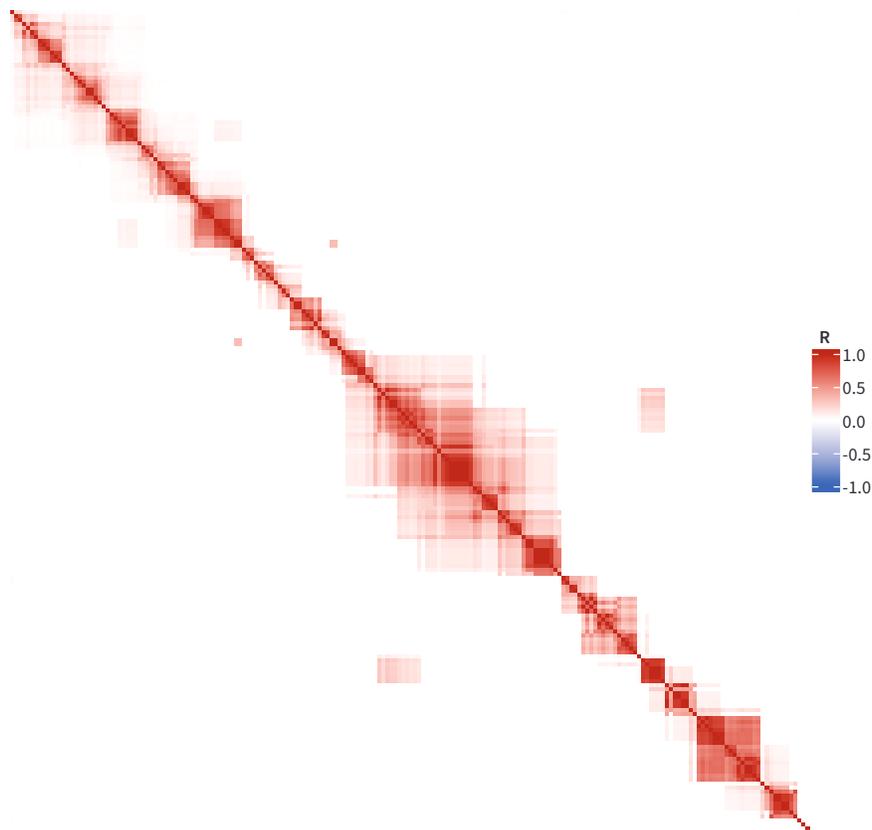


FIGURE F.7. The correlation matrix for the first 200 genes on chromosome 3 using the filtered set of samples.

395 ESTIMATING CNV LINKAGE DISEQUILIBRIUM

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

$$\mathbf{Y} = \mathbf{Z}_2 \beta_2 + \mathbf{Z}_c \boldsymbol{\lambda}_c + \boldsymbol{\epsilon}_2,$$

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

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

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

405 SIGNIFICANT AVERAGE EFFECTS

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

Trait	$\hat{\gamma}_1 \times 10^2$)	$\hat{\gamma}_1$ 95% CI	$\hat{\gamma}_2 \times 10^2$)	$\hat{\gamma}_2$ 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.

40 **LACK OF AVERAGE EFFECT IN GENOME-WIDE ASSOCIATION STUDIES**

409 For traits with significant average burden effects (Table F.1), we downloaded genome-wide association
410 study (GWAS) summary statistics from the Neale Lab (<https://www.nealelab.is/uk-biobank>, version
411 3, both sexes) based on 337,199 White British individuals in the UKB. In their processing pipeline, phe-
412 notypic values were inverse-rank normal transformed. Age, age squared, inferred sex, age-by-sex, age-
413 squared-by-sex, and 20 genotyping PCs were included as covariates.

414 We estimated conditionally independent hits and their effect sizes using GCTA-COJO [23] with param-
415 eters $-- \text{ojo-p } 5e-8 -- \text{ojo-s1 t}$. We used the genotypes of 10,000 unrelated White British individu-
416 als in the UKB to compute the LD reference panel. We polarized significant trait-associated variants into
417 ancestral and derived states with variation features obtained from the Ensembl Variation database [24].

It is made available under a [CC-BY-NC 4.0 International license](https://creativecommons.org/licenses/by-nc/4.0/).

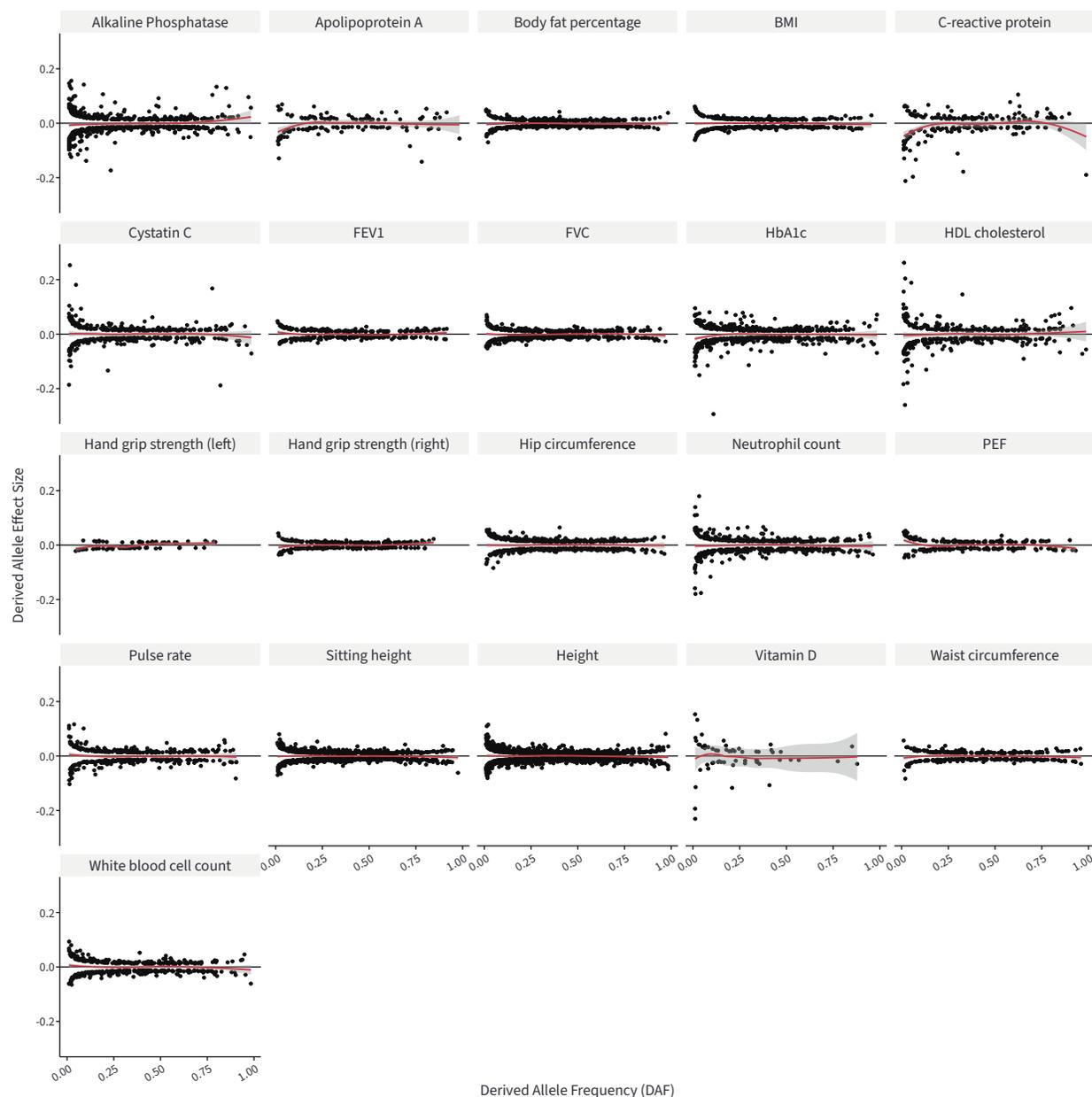


FIGURE F.8. Trumpet plots of the derived allele effect size versus the derived allele frequency for top genome-wide association study (GWAS) associations for traits with significant loss-of-function (LoF) and duplication average burden effects.

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

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

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

42 ANALYSIS OF MONOTONICITY EFFECT ESTIMATES

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

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

432 We assume that ϕ_i is drawn from a transformed beta distribution over the interval $[0, 1]$,

$$\phi_i \stackrel{\text{iid}}{\sim} \text{TransformedBeta}(\alpha, \beta, [0, 1]).$$

433 The likelihood for the transformed beta distribution can be derived using the transformation of ran-
434 dom variables approach. Suppose that $U \sim \text{Beta}(\alpha, \beta)$. We define V as

$$V = g(U) = (b - a)U + a$$

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

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

436 The inverse transform is

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

437 and the derivative is

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

43 The likelihood is therefore

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

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

440 dence,

$$\begin{aligned} \hat{\alpha}, \hat{\beta} &= \arg \max_{\alpha, \beta} \prod_{i=1}^T \mathbb{P}(\hat{\phi}_i, \alpha, \beta | \hat{c}_i) \\ &= \arg \max_{\alpha, \beta} \prod_{i=1}^T \int_0^1 \mathbb{P}(\hat{\phi}_i | \phi_i, \hat{c}_i) \mathbb{P}(\phi_i | \alpha, \beta) d\phi_i \\ &= \arg \max_{\alpha, \beta} \prod_{i=1}^T \int_0^1 \mathcal{N}(\hat{\phi}_i; \phi_i, \hat{c}_i^2) \text{TransformedBeta}(\phi_i; \alpha, \beta, [0, 1]) d\phi_i. \end{aligned}$$

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

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

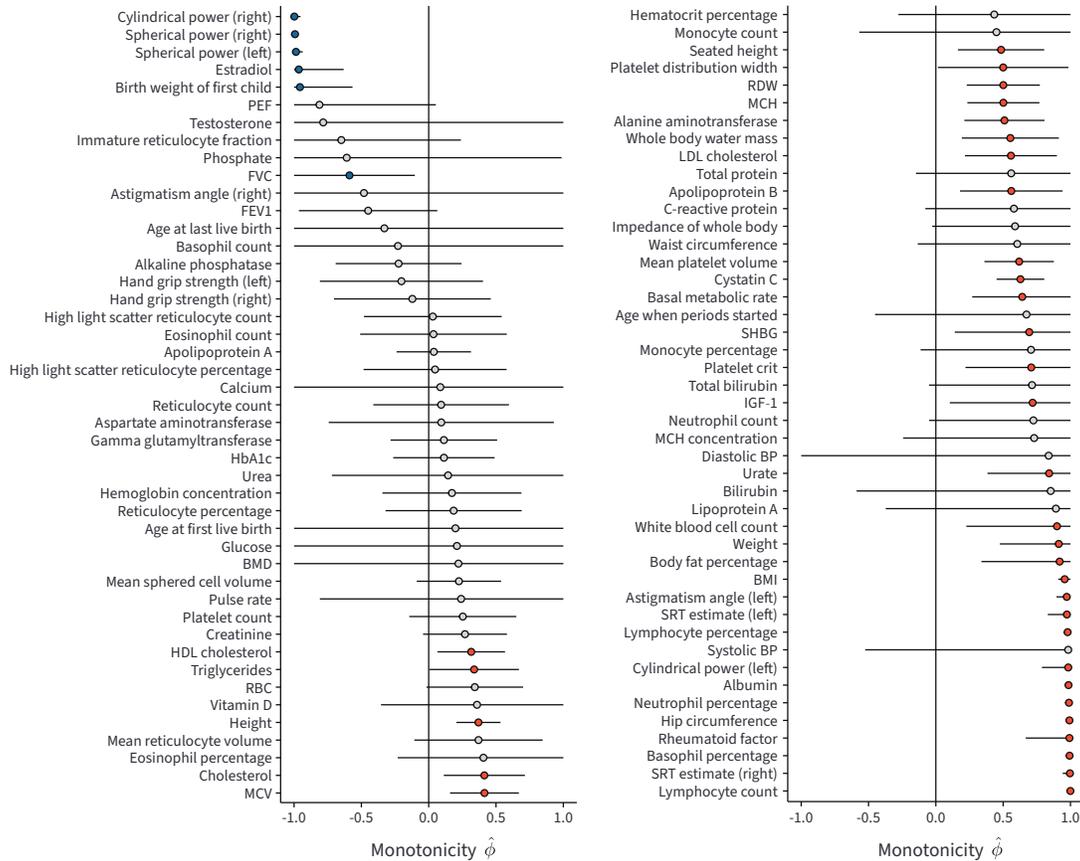


FIGURE F.9. Point estimates with 95% confidence intervals for monotonicity from the maximum likelihood estimation approach. Blue points represent significantly negative monotonicity, while red points represent significantly positive monotonicity. Grey points represent non-significant values.

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

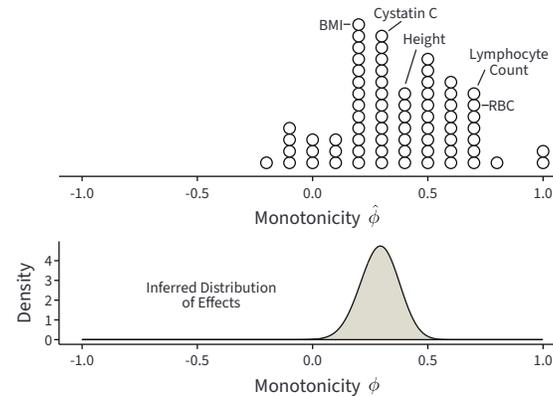


FIGURE F.10. (Top) The method-of-moments (MoM) point estimates for monotonicity for the traits that were analyzed. (Bottom) The inferred distribution of effects from a random-effects meta-analysis.

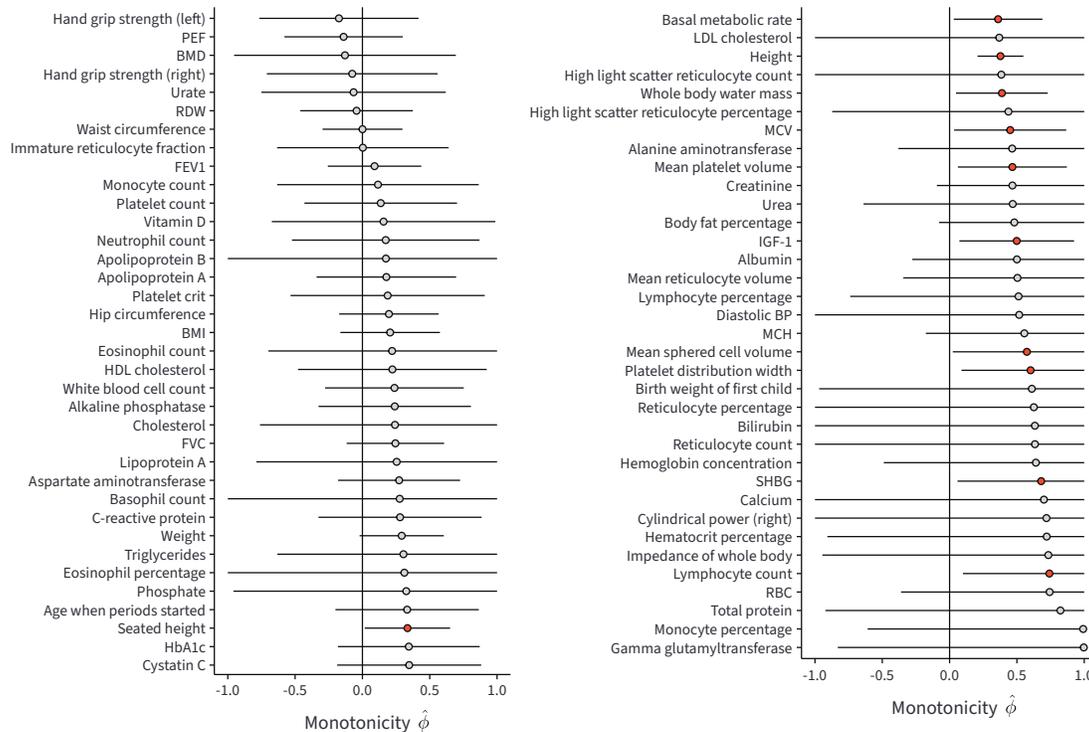


FIGURE F.11. Point estimates with 95% confidence intervals for monotonicity from the method-of-moments (MoM) approach. Red points represent significantly positive monotonicity. Grey points represent non-significant values.

452 **CONCORDANCE OF LOF AND DELETION EFFECTS**

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

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

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

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

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

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

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

$$\rho_i \stackrel{\text{iid}}{\sim} \text{TransformedBeta}(\alpha, \beta, [-1, 1]) .$$

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

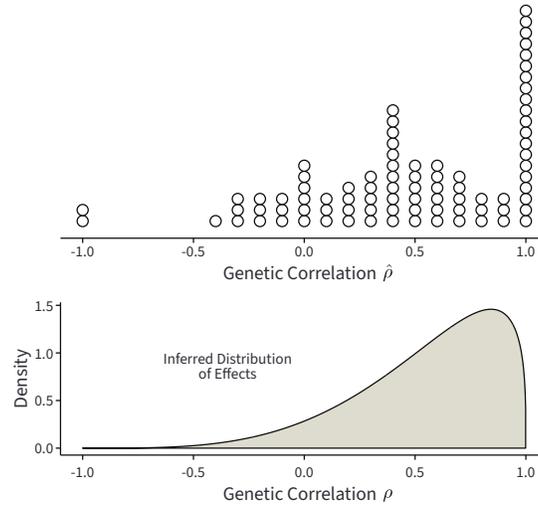


FIGURE F.12. (Top) The maximum likelihood estimates (MLEs) for genetic correlation for the traits that were analyzed. The genetic correlation is estimated between loss-of-function (LoF) burden effects and deletion burden effects. (Bottom) The inferred distribution of effects from a random-effects meta-analysis.

467 This analysis demonstrates that LoF variant and deletion burden tests are markedly similar (Figure
46 F.12). 93.53% of the density of the inferred distribution of genetic correlations is over the positive part of
469 the domain of ρ . For this analysis, we removed traits with $\hat{c}_i < 10^{-6}$ as these might represent unstable
470 estimates from the optimization procedure. The estimated genetic correlation for these estimates was
471 $\hat{\rho} = 1$, meaning that dropping them from our analysis biased our inferred distribution towards zero.

472 APPENDIX REFERENCES

- 473 [1] Chiara Auwerx et al. “The individual and global impact of copy-number variants on complex human traits”.
474 *The American Journal of Human Genetics* 109.4 (Apr. 2022), pp. 647–668. DOI: 10.1016/j.ajhg.2022.02.010.
- 475 [2] Benjamin H Good. “Linkage disequilibrium between rare mutations”. *Genetics* 220.4 (Apr. 2022), iyac004. DOI:
476 10.1093/genetics/iyac004.
- 477 [3] Xiang Zhu and Matthew Stephens. “Bayesian large-scale multiple regression with summary statistics from
47 genome-wide association studies”. *The Annals of Applied Statistics* 11.3 (Sept. 2017), pp. 1561–1592. DOI: 10.
479 1214/17-AOAS1046.
- 4 0 [4] Shun-ichi Amari. “Natural Gradient Works Efficiently in Learning”. *Neural Computation* 10.2 (Feb. 1998),
4 1 pp. 251–276. DOI: 10.1162/089976698300017746.
- 4 2 [5] Luigi Malagò and Giovanni Pistone. “Information Geometry of the Gaussian Distribution in View of Stochastic
4 3 Optimization”. *Proceedings of the 2015 ACM Conference on Foundations of Genetic Algorithms XIII*. FOGA '15.
4 4 New York, NY, USA: Association for Computing Machinery, Jan. 2015, pp. 150–162. DOI: 10.1145/2725494.
4 5 2725510.
- 4 6 [6] Larry Armijo. “Minimization of functions having Lipschitz continuous first partial derivatives.” *Pacific Journal*
4 7 *of Mathematics* 16.1 (Jan. 1966), pp. 1–3.
- 4 [7] Roy Frostig, Matthew James Johnson, and Chris Leary. “Compiling machine learning programs via high-level
4 9 tracing”. *SysML*. Stanford, CA, Feb. 2018.

- 490 [8] Charles C. Margossian. “A Review of automatic differentiation and its efficient implementation”. *WIREs Data*
491 *Mining and Knowledge Discovery* 9.4 (Nov. 2018), e1305. doi: 10.1002/WIDM.1305.
- 492 [9] Brendan K. Bulik-Sullivan et al. “LD Score regression distinguishes confounding from polygenicity in genome-
493 wide association studies”. *Nature Genetics* 47.3 (Feb. 2015), pp. 291–295. doi: 10.1038/ng.3211.
- 494 [10] Christopher C Chang et al. “Second-generation PLINK: rising to the challenge of larger and richer datasets”.
495 *GigaScience* 4.1 (Dec. 2015), s13742–015–0047–8. doi: 10.1186/s13742-015-0047-8.
- 496 [11] Daniel J. Weiner et al. “Polygenic architecture of rare coding variation across 394,783 exomes”. *Nature*
497 614.7948 (Feb. 2023), pp. 492–499. doi: 10.1038/s41586-022-05684-z.
- 49 [12] Sarah M. Urbut et al. “Flexible statistical methods for estimating and testing effects in genomic studies with
499 multiple conditions”. *Nature Genetics* 51.1 (Nov. 2018), pp. 187–195. doi: 10.1038/s41588-018-0268-8.
- 500 [13] Bernard Delyon, Marc Lavielle, and Eric Moulines. “Convergence of a Stochastic Approximation Version of
501 the EM Algorithm”. *The Annals of Statistics* 27.1 (Feb. 1999), pp. 94–128.
- 502 [14] Estelle Kuhn and Marc Lavielle. “Coupling a stochastic approximation version of EM with an MCMC proce-
503 dure”. *ESAIM: Probability and Statistics* 8 (Aug. 2004), pp. 115–131. doi: 10.1051/ps:2004007.
- 504 [15] A. P. Dempster, N. M. Laird, and D. B. Rubin. “Maximum Likelihood from Incomplete Data Via the EM Al-
505 gorithm”. *Journal of the Royal Statistical Society: Series B (Methodological)* 39.1 (Sept. 1977), pp. 1–22. doi:
506 10.1111/j.2517-6161.1977.tb01600.x.
- 507 [16] Eli Bingham et al. “Pyro: Deep Universal Probabilistic Programming”. *Journal of Machine Learning Research*
50 20.28 (Feb. 2019), pp. 1–6.
- 509 [17] Du Phan, Neeraj Pradhan, and Martin Jankowiak. *Composable Effects for Flexible and Accelerated Probabilis-
510 tic Programming in NumPyro*. Dec. 2019. doi: 10.48550/arXiv.1912.11554.
- 511 [18] Estelle Kuhn, Catherine Matias, and Tabea Rebafka. “Properties of the stochastic approximation EM algo-
512 rithm with mini-batch sampling”. *Statistics and Computing* 30.6 (Nov. 2020), pp. 1725–1739. doi: 10.1007/
513 s11222-020-09968-0.
- 514 [19] Joshua D. Backman et al. “Exome sequencing and analysis of 454,787 UK Biobank participants”. *Nature*
515 599.7886 (Oct. 2021), pp. 628–634. doi: 10.1038/s41586-021-04103-z.
- 516 [20] Nasa Sinnott-Armstrong et al. “Genetics of 35 blood and urine biomarkers in the UK Biobank”. *Nature Genetics*
517 53.2 (Jan. 2021), pp. 185–194. doi: 10.1038/s41588-020-00757-z.
- 51 [21] Tony Zeng et al. “Bayesian estimation of gene constraint from an evolutionary model with gene features”.
519 *Nature Genetics* (July 2024), pp. 1–12. doi: 10.1038/s41588-024-01820-9.
- 520 [22] Shuwei Li et al. *Whole-genome sequencing of half-a-million UK Biobank participants*. Dec. 2023. doi: 10.1101/
521 2023.12.06.23299426.
- 522 [23] Jian Yang et al. “Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies addi-
523 tional variants influencing complex traits”. *Nature Genetics* 44.4 (Apr. 2012), pp. 369–375. doi: 10.1038/ng.
524 2213.
- 525 [24] Fergal J Martin et al. “Ensembl 2023”. *Nucleic Acids Research* 51.D1 (Jan. 2023), pp. D933–D941. doi: 10.1093/
526 nar/gkac958.
- 527 [25] Roshni A. Patel et al. *Conditional frequency spectra as a tool for studying selection on complex traits in
52 biobanks*. June 2024. doi: 10.1101/2024.06.15.599126.
- 529 [26] Brendan Bulik-Sullivan et al. “An atlas of genetic correlations across human diseases and traits”. *Nature Ge-
530 netics* 47.11 (Sept. 2015), pp. 1236–1241. doi: 10.1038/ng.3406.