Factorizing polygenic epistasis improves prediction and uncovers biological pathways in complex traits

David Tang[1,2,*], Jerome Freudenberg[1], and Andy Dahl[1,**]

[1]Section of Genetic Medicine, University of Chicago
[2]Program in Bioinformatics and Integrative Genomics, Harvard Medical School
[*]davidtang@g.harvard.edu
[**]andywdahl@uchicago.edu

Abstract

Epistasis is central in many domains of biology, but it has not yet proven useful for complex traits. This is partly because complex trait epistasis involves polygenic interactions that are poorly captured in current models. To address this gap, we develop a new model called Epistasis Factor Analysis (EFA). EFA assumes that polygenic epistasis can be factorized into interactions between a few latent pathways, or Epistasis Factors (EFs). We mathematically characterize EFA and use simulations to show that EFA outperforms current epistasis models when its assumptions approximately hold. Applied to predicting yeast growth traits, EFA outperforms the additive model for several traits with large epistasis heritability and uniformly outperforms the standard epistasis model, which we replicate in a second dataset. We also apply EFA to four traits in the UK Biobank and find statistically significant evidence for epistasis in male testosterone. Moreover, we find that the inferred EFs partly recover known biological pathways. Our results demonstrate that realistic statistical models can identify meaningful epistasis in complex traits, indicating that epistatic models have promise for precision medicine and characterizing the biology underlying GWAS results.

Introduction

Epistasis refers to interactions between genetic effects on a trait. Epistasis is central in many domains of biology, including rare human disorders [1–3], protein evolution [4, 5], natural selection [6, 7], and functional genomics [8]. Statistical models of epistasis can be useful for characterizing genetic architecture [9–16], improving genomic selection [17, 18], and unbiasedly screening for unknown genomic mechanisms [19–22]. In striking contrast, it remains debated if epistasis matters in complex traits [23–25]. This is primarily because complex trait biology remains poorly understood, severely limiting experimental characterization of complex trait epistasis.

Unbiased genome-wide statistical tests for epistasis could fill this gap in our understanding of complex traits. However, modelling epistasis in complex traits is challenging because they are affected by a large number of genetic variants (**Figure 1a**). Although studies have identified some examples of epistasis in complex traits [26–28], they have not explained significant missing heritability [29, 30] nor shed much light on complex trait biology. We hypothesize this is partly due to shortcomings in current models of complex trait epistasis. Specifically, prior studies have used an unrealistic and under-powered model that assumes each epistatic interaction effect is completely independent of additive effects and other interaction effects (**Figure 1b**, [31–33]).

Here, we develop a new complex trait epistasis model called Epistasis Factor Analysis (EFA). EFA assumes a "coordinated" form of epistasis where SNP-SNP interactions are structured by interactions between a few latent pathways, which we call Epistasis Factors (EFs) (**Figure 1c**, [33, 34]). EFA simultaneously estimates EFs and their interactions solely from GWAS data. The inferred EFs can be validated with known genomic annotations and used to discover novel biological pathways underlying GWAS results. Additionally, EFA can improve prediction over additive models, which we demonstrate in simulations and two multi-trait yeast datasets. Finally, we show that EFA has power to detect statistically significant epistasis in complex human traits and, furthermore, that the inferred EFs partly recover known biological pathways. Our work shows that epistasis can matter in complex traits if we use more realistic models.

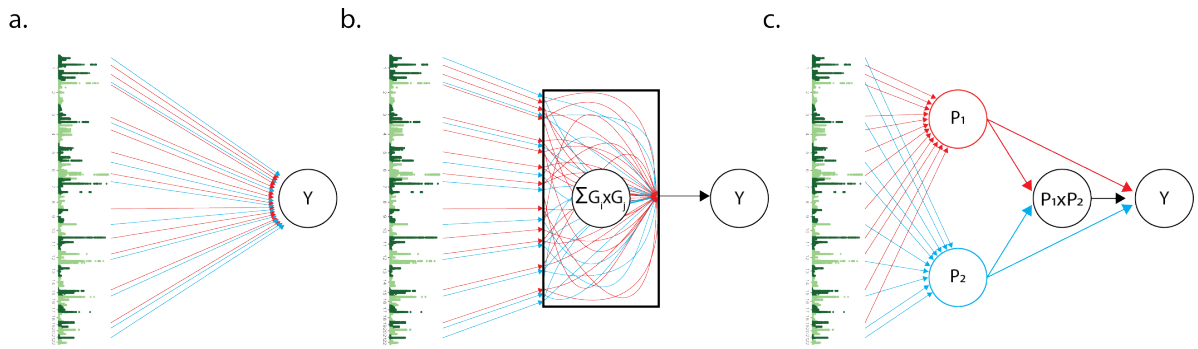

Figure 1: Three genetic architectures consistent with a typical GWAS. (a) Additive: Each SNP has an independent effect on the phenotype, Y. (b) Uncoordinated: Each SNP interacts randomly with all other loci. (c) Coordinated: SNPs can be grouped into a few interacting pathways. EFA uses the P1xP2 interaction to improve statistical power to detect epistasis and to sort GWAS loci into biologically meaningful pathways.

Epistasis Factor Analysis

Nearly all complex trait studies assume the additive model, and it remains debated whether epistasis matters in complex traits [24, 25]. However, this debate has largely been predicated on an "uncoordinated" model of epistasis that assumes the epistasis effects are completely random–independent of both additive effects and other epistasis effects [30, 31, 35]. Uncoordinated epistasis is not interpretable, realistic, or statistically supported [30]. In contrast, we recently proposed a more realistic "coordinated" model of epistasis where epistasis and additive effects are structured by latent pathways (Figure 1c, [33]). Coordination is motivated by known forms of epistasis in simpler traits, such as genetic modifiers of a Mendelian gene [2].

Here, we introduce a new coordinated epistasis model to identify these latent pathways that we call Epistasis Factor Analysis (EFA). EFA makes two key assumptions on the nature of epistasis in complex traits:

A1: SNP effects are mediated through a few distinct pathways

A2: These pathways interact

A1 and A2 are biologically motivated. For example, the pathways in A1 could reflect distinct causal tissues [36, 37], core genes [38, 39], or heritable exposures like smoking [40]. In accordance with A2, causal tissues might signal each other, core genes might encode proteins that physically interact, or an exposure's effect may depend on underlying genetic risk.

The key idea in EFA is to share information between additive effects and epistasis effects. On one hand, leveraging additive signals greatly improves power to detect epistasis. On the other hand, leveraging epistasis enables unbiasedly partitioning SNPs into pathways, which is impossible in the additive model without external data or biological annotations.

Concretely, the EFA model is:

$$y_i = P_{i1} + P_{i2} + \lambda P_{i1} P_{i2} + \varepsilon_i; \qquad P_{ik} := \sum_j G_{ij} U_{jk} \tag{1}$$

where P_{ik} is the level of EF k for sample i, U_{jk} is the weight of SNP j on EF k, and λ is the interaction between EFs. We call this Epistasis Factor Analysis because it implicitly makes a low-rank factorization to the genome-wide epistasis matrix (**Methods**). We generalize to $K > 2$ EFs in the **Methods**, and we theoretically characterize EFA in the **Supplement**. In practice, G will typically contain dozens of SNPs that have been pre-screened based on their additive effects. We fit U and λ as fixed effects using maximum likelihood (**Methods**).

EFA recovers latent pathways and improves prediction in simulations

We performed simulations to characterize EFA as a function of additive heritability (h^2_{add}), epistasis heritability (h^2_{epi}), and the fraction of epistasis heritability due to coordinated vs uncoordinated effects (**Methods**).

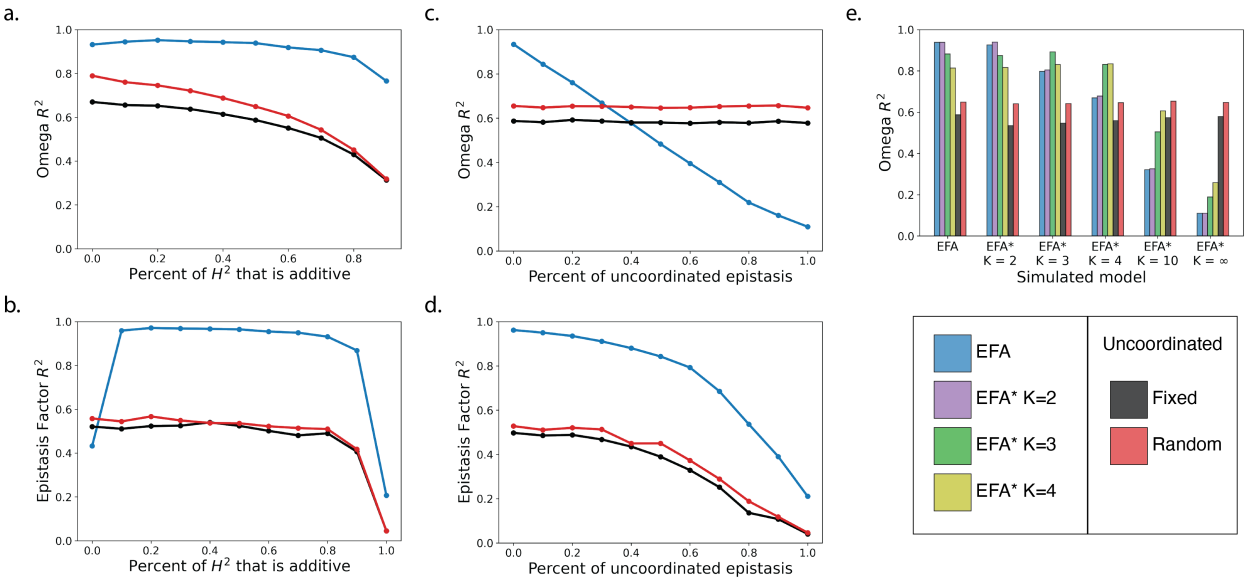

Figure 2: Simulations characterize the accuracy of polygenic epistasis estimates. Accuracy is quantified by the squared correlation between estimated and true pairwise SNP epistasis effects, ω (a,c,e) or between estimated and true EFs, U (b,d). Simulations under the EFA model vary the fraction of heritability due to additive vs epistasis effects (a,b). Partial coordination is simulated by combining epistasis effects from the EFA model with uncoordinated epistasis effects (c,d). Other forms of coordination are simulated using the EFA* model with $K \geqslant 2$ (e), which is a more general version of EFA (**Methods**).

We define broad-sense heritability as $H^2 := h^2_{add} + h^2_{epi}$, and always use $H^2 = 0.5$. We compared EFA to uncoordinated epistasis estimates fit either by fixed or random effects (**Methods**). We evaluated estimation accuracy for the pairwise SNP-SNP interactions, $\omega_{jj'}$ (see (3) in the **Methods**), and for the EFs, U_{jk}. EFA directly fits EFs and implicitly fits ω (**Methods**). Uncoordinated methods directly fit ω, and we derive uncoordinated EFs *post hoc* using PCs of the estimated ω matrix (**Methods**).

We first simulated from the EFA model in (1) with $N = 1,000$ samples and $M = 20$ SNPs. As expected, EFA outperforms uncoordinated estimates of ω across the range of h^2_{epi} (**Figure 2a**). EFA's gain is greatest when $h^2_{add} \gg h^2_{epi}$, illustrating how EFA's epistasis estimates borrow strength from additive effects. Analogously, EFA improves additive effect estimates over standard additive models when $h^2_{epi} \gg h^2_{add}$ (**Supplementary Figure 1**), though this scenario is not likely in practice. EFA also substantially outperforms uncoordinated estimates of EFs (**Figure 2b**); however, all methods perform similarly when $h^2_{add} \approx 0$ or $h^2_{epi} \approx 0$, because coordination essentially vanishes in these situations.

We next fix $h^2_{add} = h^2_{epi} = 25\%$ and vary the level of coordination by adding in uncoordinated epistasis effects (**Methods**). As expected, EFA's performance decays as coordination weakens, while the uncoordinated estimates remain stable (**Figure 2c**). Nonetheless, EFA always outperforms uncoordinated estimates of the EFs because EFA specifically targets the coordinated component of epistasis (**Figure 2d**).

We next simulated and fit $K > 2$ EFs using a generalization of EFA called EFA* (**Methods**). As expected, ω estimates were more accurate when fitting EFA* with the true value of K (**Figure 2e**), though the base EFA model was robust to modest misspecification of K. However, the uncoordinated model performed better for $K = 10$ and much better for $K = \infty$, which is equivalent to uncoordinated epistasis [33]. We did not consider EF estimation accuracy because EFs are not identified for $K > 2$ (**Supplement**).

We next varied the number of SNPs, M, and sample size, N (**Supplementary Figure 3**). Across all N and M, EFA always outperforms uncoordinated estimates. Performance for all methods decays as M grows because each individual SNP effect is smaller. This illustrates why modelling epistasis in complex traits is difficult. Likewise, all methods improve as N grows.

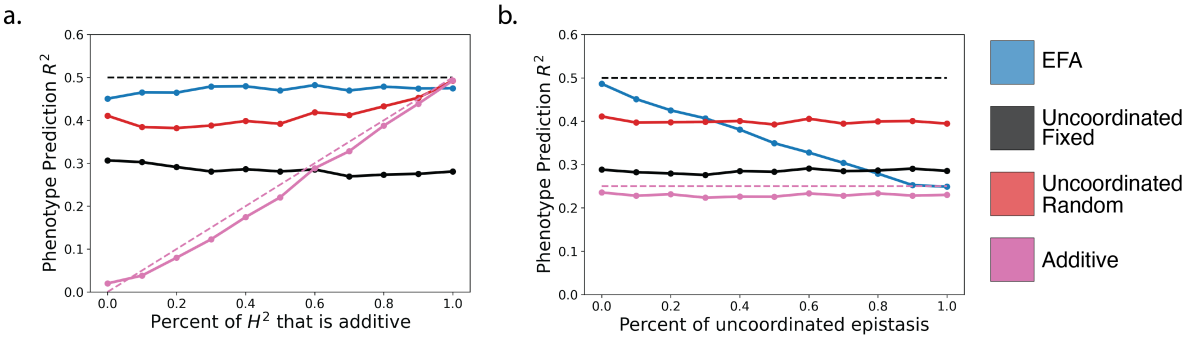

Figure 3: Phenotype prediction accuracy in simulations. (a) varies the fraction of heritability due to additive vs epistasis effects. (b) varies the fraction of epistasis effects due to coordinated pathway-level interactions vs uncoordinated random interactions.

Finally, we asked if EFA can improve phenotype prediction. We first simulated from EFA's model and found that EFA performed well, with prediction R^2 near the theoretical limit of H^2 (**Figure 3a**). EFA always outperformed the fixed-effect uncoordinated model, and EFA outperformed the random effect uncoordinated model except when epistasis was absent. (We derive the predictions from the uncoordinated random effect model in the **Supplement**.) We then fixed $h^2_{epi} = h^2_{add} = .25$ and varied the fraction of epistasis due to coordination, as above. Results were similar to ω estimation accuracy: when epistasis is mostly coordinated, EFA is optimal, but EFA performs poorly when epistasis is uncoordinated (**Figure 3b**). Overall, EFA improves estimation of latent pathways in all our tests and improves prediction when epistasis is sufficiently strong and coordinated.

EFA improves prediction in complex yeast traits

We next asked if EFA could improve genetic prediction in complex traits. We studied 46 complex yeast traits measured on 1008 samples, where each trait was growth rate on a different medium [13]. These traits have varying levels of epistasis heritability, as defined by the difference between the phenotypic similarity of clones (H^2) and the additive heritability estimated from genotype data (h^2). We measured prediction accuracy by the squared correlation between predicted and observed phenotypes using 10-fold cross-validation. For each trait, this yielded 30-70 clumped+thresholded SNPs ($r^2 < 0.2$, **Methods**).

We first compared EFA to the additive model and found that EFA significantly improved prediction accuracy for 6/46 traits ($p < .05$, paired t-test, **Methods**, **Supplementary Table 1**). These 6 traits all had large epistasis heritability (**Figure 4a**). More generally, the difference between R^2_{EFA} and R^2_{add} correlated with the epistasis heritability (Spearman's $\rho = 0.48$, $p < 0.001$). This is consistent with our simulations showing that EFA can outperform additive predictions when h^2_{epi} is high. We next compared EFA to uncoordinated epistasis predictions using fixed effects and found that EFA was superior for all 46 traits, regardless the level of h^2 and H^2 (**Figure 4b**).

Next, we compared EFA to additive and epistasis random effects models. Unlike fixed effect models, including EFA, random effect models can fit the whole genome without LD pruning, which is often a major advantage for prediction in complex traits [18, 42, 43]. Nonetheless, EFA significantly outperforms both additive and uncoordinated epistasis random effect models for 7/46 traits ($p < .05$). We also compared the two random effect models, and we found that the uncoordinated epistasis model significantly outperformed the additive model for 9 traits ($p < .05$). Intriguingly, these 9 traits only overlap 1 of the 7 traits where EFA outperforms the additive random effect model, suggesting that the degree of coordination varies across traits.

Finally, we evaluated replication in a second dataset from the same group with 4390 samples measured on 20 of the above 46 traits [14]. We replicated the superiority of EFA over the additive model for all 3 traits where EFA beat the additive model previously ($p < .05$; the other 3/6 traits were not measured in the second dataset). Further, at this larger sample size, we find 4 additional traits where EFA significantly outperforms

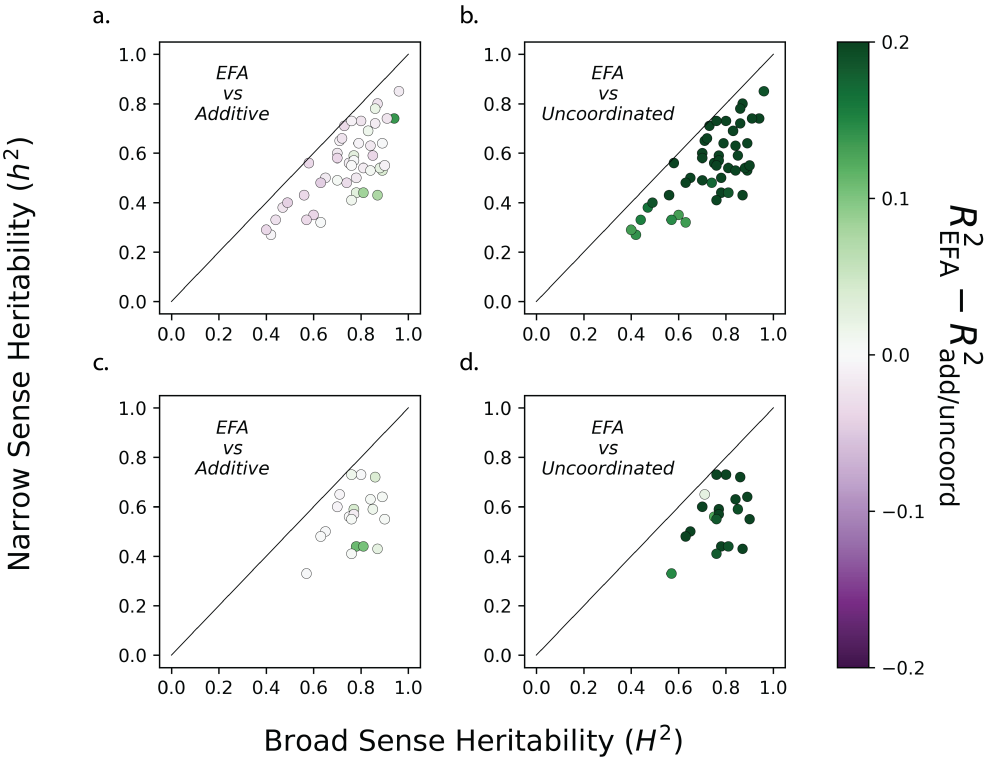

Figure 4: Prediction accuracy in complex yeast traits. Each point is a trait with narrow- and broad-sense heritability estimates from [41]. (a) EFA significantly outperforms the additive model for 6/46 traits from Bloom et al 2013 data. (c) EFA significantly outperforms the additive model for 7/20 traits from Bloom et al 2015 data. (b,d) EFA always outperforms the uncoordinated fixed-effect epistasis model.

the additive fixed effect model (**Figure 4c**, **Supplementary Table 2**). We also replicate that EFA always outperforms the uncoordinated fixed effects model (**Figure 4d**). Overall, EFA improves prediction over the additive model for some complex traits, and the utility of EFA will likely grow with larger sample sizes.

EFA detects epistasis and uncovers biological pathways in complex human traits

We applied EFA to complex human traits in UK Biobank. We studied the four traits with extensively characterized GWAS in [44]: urate, insulin-like growth factor 1 (IGF1), and testosterone in males and females. We used the top 100 LD-pruned GWAS hits for each trait (all 77 for female testosterone, **Methods**).

We first asked if EFA could detect statistically significant epistasis. Because EFA is a complex latent variable model, it is not straightforward to implement traditional hypothesis tests, such as likelihood ratio tests. Instead, we use a new nonparametric test based on the R^2 between SNP-SNP epistasis estimates (ω) from independent halves of the samples. We test significance using an empirical null distribution that is generated by permuting phenotypes (**Methods**). We show that this is a calibrated test of epistasis by performing simulations under the additive model (**Supplementary Figure 6**). We also apply this same nonparametric testing approach to test uncoordinated epistasis estimates and to Epistasis Factors.

EFA identified significant epistasis in male testosterone ($p = 0.022$, **Figure 5a**). The EFs for male testosterone were also significant ($p = 0.040$, **Supplementary Figure 7**). By contrast, the uncoordinated epistasis estimates were not significant in male testosterone ($p > 0.05$, **Supplementary Figure 7**), consistent with prior studies that did not identify epistasis in this trait [44]. In urate and IGF1, by contrast, the uncoordinated estimates were significant ($p = 0.004$, 0.017) while EFA's estimates were not (**Supplementary Figure 7**). This is surprising because a prior study did not detect uncoordinated epistasis in these traits

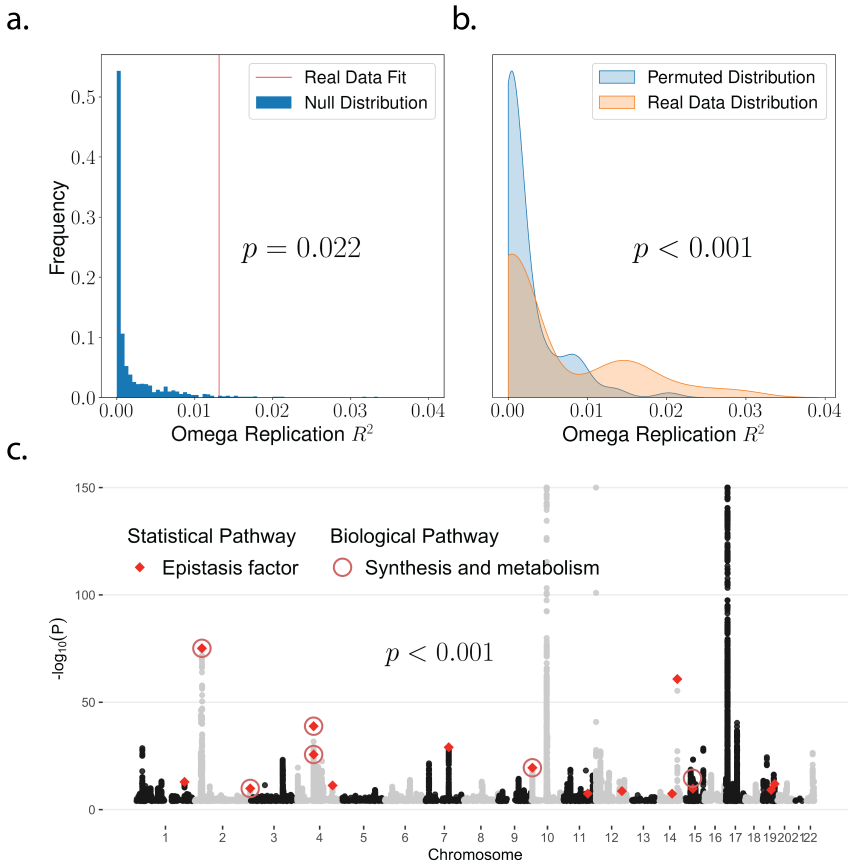

Figure 5: EFA identifies statistically and biologically significant epistasis in male testosterone. (a) EFA pathways replicate between two random halves of the UKB male individuals in testosterone. (b) Across 100 random sample splits, EFA replicates better than expected by chance based on permuted phenotypes. (c) Steroid synthesis and metabolism genes (defined by [44] are enriched in EF1 in male testosterone.

[30], though that study used a different test and a different model (genome-wide random effects).

We next combined evidence across 100 random 50-50 splits (**Methods**). This improved power and gave even stronger support for EFA in male testosterone ($p = 0.0003$, paired t-test, **Figure 5b**, **Methods**). The p-values for the EFs also grew stronger ($p = 0.012$, **Supplementary Figure 10**). The uncoordinated model remained insignificant (**Supplementary Figure 11**).

We next assessed robustness to dominance and LD using simulations. First, we simulated a dominance model without any epistasis between distinct SNPs (**Supplement**, **Supplementary Figure 6**). Under strong dominance ($h^2_{dom} = 1\%$, [30, 45]), the uncoordinated model is about 10-fold inflated (FPR=57% at nominal 5% level) and EFA is about 3-fold inflated (FPR=16%). Both are expected: EFA does not assume dominance is absent (unlike our prior Even-Odd test [33]), but it does emphasize inter-SNP epistasis more than the uncoordinated model. Under more realistic levels of dominance ($h^2_{dom} = 0.1\%$), both tests are roughly calibrated (**Supplementary Figure 6**). Second, we simulated an additive model with unobserved causal variants in LD with measured variants. Our simulation meets the necessary condition for 'phantom' epistasis in [46] (**Supplement**). We do not observe any bias from LD in EFA (**Supplementary Figure 6**). Nonetheless, other forms of LD could potentially cause bias in EFA. Overall, the polygenic nature of EFA mitigates bias from LD and dominance, which can severely bias SNP-level epistasis tests [18, 47–50].

As epistasis can be scale-dependent [51], we next tested robustness by quantile-normalizing the phenotypes (**Supplementary Figure 9**). This eliminated the significance of EFA for male testosterone, though urate

and female testosterone became significant. These discrepancies could reflect either false positives or false negatives from overly-conservative scaling [33].

We next asked if the male testosterone EFs capture meaningful biology by comparing to the known biological pathways defined in [44]. Specifically, we used a hypergeometric test for enrichment of the top 15 SNPs per EF in each predefined biological pathway (relative to the other 85 GWAS SNPs, **Table 1**). EF1 is strongly enriched in "steroid synthesis and metabolism" ($p = 0.0002$). This enrichment is driven by SNPs assigned to the genes *UGT1A8*, *SRD5A2*, *UGT2B15*, *UGT2B17*, and *AKR1C4*, which live on different chromosomes (2, 4, and 10). We confirmed that this enrichment is robust if we instead use the top 10 or 20 SNPs per EF (**Supplementary Table 3**). The results are also robust if we increase the window size used to assign SNPs to genes, or if we further prune SNPs such that none are assigned to the same gene ($p = 0.004$, **Supplementary Table 7**).

Trait	Biological Pathway	# SNPs in EF1	# SNPs in EF2	# SNPs total
IGF-1	Downstream signaling	1 ($p = 0.403$)	0 ($p = 1.000$)	3
	Growth horm. secr.	2 ($p = 0.114$)	1 ($p = 0.499$)	4
	IGF-1 secretion	1 ($p = 0.499$)	1 ($p = 0.499$)	4
	IGF-1 serum balance	0 ($p = 1.000$)	3 ($p = 0.074$)	7
	Ras signaling	0 ($p = 1.000$)	1 ($p = 0.581$)	5
Urate	Solute transport	2 ($p = 0.710$)	6 ($p = 0.010$) ⋆	15
	Purine metabolism	0 ($p = 1.000$)	0 ($p = 1.000$)	2
Testosterone (male)	Steroid synthesis & metabolism	5 ($p < 0.001$) ⋆	0 ($p = 1.000$)	6
	HPG signaling	1 ($p = 0.564$)	0 ($p = 1.000$)	5
	Serum homeostasis	0 ($p = 1.000$)	0 ($p = 1.000$)	3
Testosterone (female)	Serum homeostasis	0 ($p = 1.000$)	0 ($p = 1.000$)	1
	Steroid synthesis & metabolism	1 ($p = 0.795$)	0 ($p = 1.000$)	7

Table 1: Epistasis Factors are enriched in predefined biological pathways from [44]. p-values are from a hypergeometric test of the top 15 EF SNPs in each biological pathway.

We next tested pathway enrichment in the other three traits. We found that EF2 in urate is enriched in SNPs associated with "solute transport" ($p = 0.010$, **Table 1**). However, this primarily reflects a single locus with large effect (*SLC2A9*, **Supplementary Figure 12**). In particular, restricting to one SNP per gene eliminates this enrichment ($p = 0.89$). While this EF signal for urate could reflect biologically meaningful *cis-trans* epistasis [33, 39], it could also merely reflect subtle LD with unmeasured additive and/or dominance effects [30]. Altogether, EFA is capable of finding biologically meaningful epistasis within and between loci, but the former is more susceptible to statistical false positives.

Discussion

Despite strong evidence for epistasis in other domains of biology, it remains unclear if epistasis is important in complex traits. We have introduced a model to address this gap called Epistasis Factor Analysis (EFA). In contrast to existing uncoordinated models of polygenic epistasis, which lack biological motivation, EFA assumes a structured form of polygenic epistasis implied by pathway-level interactions (**Figure 1C**). We showed that EFA outperforms additive predictions in several yeast traits, which we replicated in a second dataset. Additionally, we found that EFA can identify epistasis in complex human traits where standard models fail and, furthermore, that EFA can learn Epistasis Factors which unbiasedly recover known biological pathways. Broadly, these successes show that epistasis has promise for characterizing the biology underlying GWAS hits, which remain largely inscrutable.

Several other polygenic epistasis models have been developed for complex traits. The best-studied models, which we call "uncoordinated" [33], assume that all SNP-SNP epistasis effects are independent of each other and of additive effects [31]. While uncoordinated models are mathematically rigorous, they are biologically implausible and statistically under-powered [30]. Nonetheless, one surprise in our study was the success of uncoordinated models in both yeast prediction and complex human traits. We attribute the former to our use of random effect predictions, as the fixed effect predictions are uniformly poor. The latter is due

either to our novel nonparametric test or our restriction to GWAS loci [24, 52], which implicitly assumes a connection between additive and epistasis effects. These results show the importance of complex trait epistasis completely independently of our EFA model.

More recent polygenic epistasis models make simplifying assumptions to improve power and interpretation. For example, autosome-sex interactions represent a particular form of epistasis that strongly affects many complex traits [53–56], and interactions between a locus (or region) with a polygenic background can identify epistasis hubs [57–62]. EFA instead assumes that polygenic epistasis is driven by interactions among a few latent polygenic pathways. We and others have previously argued for this latent pathway model [33, 34], but EFA is the first approach to explicitly identify these latent pathways. BANN [63] is conceptually similar to EFA in that it models complex traits as epistatic compositions of simpler components. In contrast, the EFA pathways are fewer, more polygenic, and do not require prior biological annotations.

There are several limitations to our approach. First, EFA does not scale to genome-wide genotypes. In practice, we pre-screen SNPs based on additive signals, which enriches for SNPs with epistasis effects [24, 52]. Second, our epistasis tests are polygenic and thus do not establish epistasis for any given SNP. In the future, sparse models (Bayesian or frequentist) could test which SNPs affect which EFs. Third, statistical interaction tests are susceptible to false positives from phenotype scaling [33, 51, 64, 65] and/or LD with unmeasured causal variants [18, 47–50]. Indeed, the EFA signals in UKB depend on phenotype scale, and the biological EF enrichment in urate is driven by a single large-effect gene. Nonetheless, coordination is robust to modest rescaling [33] and we always prune SNPs in LD. Furthermore, these caveats do not affect our conclusion that EFA can add value for prediction and biological characterization beyond additive and uncoordinated models.

There are several potential extensions to EFA that may prove useful. First, EFA could jointly model multiple traits that partly share latent pathways. This would add power to detect shared pathways and provide a richer model of pleiotropy than genetic correlation, which is overly simplistic [66–70]. Second, EFA could be extended to complex disorders using a generalized linear model framework, which could identify limiting disease pathways [34] and/or etiologically distinct disease subtypes [55]. However, binary traits will have lower power to detect epistasis and more severe scale dependence [71]. Third, we have used SNPs in G, but more powerful and/or interpretable genetic features could be used, such as imputed gene expression [72, 73], copy number variants [27, 28, 74], or polygenic scores for secondary traits [75] or exposures [40]. EFA could also incorporate non-genetic variables such as epigenomic marks, medical image-derived features [76], disease symptoms, or electronic health records. Ultimately, we hope that the EFA model and its strong results in practice will provide solid groundwork for further study of epistasis in complex traits.

Code availability

A python implementation of EFA is available at: `https://github.com/tangdavid/efa`. This link also contains the code for all analyses in this paper, though the UKB analysis is not fully reproducible because the raw data is not public.

Data availability

We downloaded the first yeast data set from `http://genomics-pubs.princeton.edu/YeastCross_BYxRM/data/cross.Rdata` and the second yeast dataset from `https://static-content.springer.com/esm/art%3A10.1038%2Fncomms9712/MediaObjects/41467_2015_BFncomms9712_MOESM729_ESM.zip`

This research has been conducted using the UK Biobank Resource under application number 30397.

Author contributions

D.T. developed statistical methodology, performed analysis, and wrote the manuscript. J.F. performed analysis. A.D. conceived and supervised the project and wrote the manuscript.

Acknowledgements

We thank Matthew Stephens, Sasha Gusev, Sriram Sankararaman, and Noah Zaitlen for helpful feedback. We also thank the participants in UKB for making this study possible. Finally, we thank the Center for Research Informatics and the Research Computing Center for providing the compute resources necessary for carrying out this project. A.D. is supported by K25HL157603.

Methods

Mathematical description of Epistasis Factor Analysis

Epistasis Factor Analysis (EFA) aims to find a low-dimensional representation of polygenic epistasis that is both statistically powerful and biologically interpretable. The EFA model is:

$$y = G\,U\,\mathbf{1}_K + ((GU) \bullet (GU))\,\mathrm{vec}(\Lambda) + \varepsilon, \tag{2}$$

where y is a quantitative trait measured on N samples, G is an $N \times M$ matrix of genotypes at M SNPs, U is an $M \times K$ matrix of SNP effects on each of K pathways, Λ is a symmetric $K \times K$ matrix of pathway-level interactions, $\bullet$ is the face-splitting product ($A \bullet B$ is a matrix where each column is an element-wise product of a column of A and a column of B), and ε is independent and identically distributed (i.i.d.) Gaussian noise.

Intuitively, each column of U represents the SNP weights on a different latent pathway (or EF), and $P_{,k} := GU_{,k}$ is an individual's genetic level of the k-th EF. $\Lambda_{kk'}$ represents the pathway-level interaction between EFs k and k'. When $\Lambda_{kk'} > 0$, pathways k and k' interact synergistically, meaning that the magnitude of their combined phenotypic effect exceeds the sum of their individual effects. The reverse is true for antagonistic pathways where $\Lambda_{kk'} < 0$.

In the main text, we call this general model EFA* and focus instead on the special case where (i) $K = 2$ and (ii) $\Lambda_{11} = \Lambda_{22} = 0$. We focus on this special case for simplicity and also because it is identified, i.e., the estimated EFs and their interactions are meaningful. In contrast, the general EFA* model in (2) is not identified without further assumptions, such as that U is sparse. We prove these facts and comprehensively characterize the theoretical properties of EFA in the **Supplement**. We also note that most factor models, including PCA, are not identified without additional assumptions. We also calculate the level of coordinated Epistasis under EFA in the **Supplement** [33].

EFA is a special case of the standard pairwise polygenic epistasis model that underlies many epistasis models for complex traits:

$$y_i = \sum_j G_{ij}\beta_j + \sum_{j,j'} G_{ij}G_{ij'}\omega_{jj'} + \varepsilon_i \tag{3}$$

where β_j is the additive effect of SNP j, and $\omega_{jj'}$ is the epistatic interaction between SNPs j and j'. The pathway-level interaction model in EFA can be directly connected with this SNP-level epistasis model by the linked factorization of the SNP-level additive and epistasis effects:

$$\beta = \sum_k U_{,k}; \quad \omega = U\Lambda U^T$$

We estimate the parameters of EFA using maximum likelihood. We numerically optimize the log-likelihood with gradient decent based on automatic differentiation. We use multiple random initializations of the optimization to account for non-convexity in the likelihood.

Uncoordinated models of polygenic epistasis

We compare EFA with other polygenic models in simulations and real data analyses. All comparison methods are special cases of the overarching polygenic pairwise epistasis model in (3).

First, we compare to the standard additive model, which ignores epistasis by setting $\omega_{jj'} = 0$ for all j, j'. Because we mostly study scenarios with $M < N$, we primarily fit the additive model with fixed effects (by

ordinary least squares). The predictions from this model are then essentially Polygenic Risk Scores (PRS). We also fit the additive model with random effects (assuming each β_j is i.i.d. Gaussian) in the real yeast analysis, enabling us to fit the full genome-wide data without any pruning.

Second, we compare to the standard pairwise epistasis model in (3) fit with fixed-effects. Because this model has $O(M^2)$ parameters, jointly fitting all pairwise epistasis effects is often noisy or even impossible. Instead, we fit each pairwise interaction $\omega_{jj'}$ one-at-a-time (for $j \leqslant j'$) while controlling for additive fixed effects (note that only $\omega_{jj'} + \omega_{j'j}$ is identified).

Third, we compare to the uncoordinated random effect model for pairwise epistasis [31]. This model assumes that $\omega_{jj'}$ are drawn i.i.d. Gaussian. This assumption simplifies computation because the ω can be easily marginalized out of the likelihood. We fit the variance components (one for β and one for ω) with maximum likelihood. We derive the best linear unbiased predictors (BLUPs) for both ω and for phenotype prediction in the **Supplement**.

Finally, we also develop a novel approach to estimate latent pathways from uncoordinated estimates by applying PCA to the estimated ω matrix. This is a naive baseline approach that only involves post-processing uncoordinated estiamtes. In contrast, EFA directly learns pathways, linking the low dimensional components of ω with the additive effects.

Simulations

We use a polygenic simulation framework that is fully described in the **Supplement**. In brief, we simulate under the EFA model (2) with independent SNPs. The latent pathways, $U_{,k}$, are drawn from i.i.d. Gaussians with variance chosen to give the appropriate additive heritability (recall that the additive effect is related to the latent pathways by $\beta = \sum_k U_{,k}$). The upper triangular entries of Λ are also drawn fro i.i.d. Gaussians with variance chosen to give the appropriate epistasis heritability. The pairwise polygenic epistasis effects are then generated by $\omega = \sum_{k,k'} \lambda_{kk'} U_{,k} U_{,k'}^T$.

We also simulate partly coordinated epistasis by adding i.i.d. Gaussian random variables to each entry of a coordinated ω. By varying the relative contribution of these effects, we can interpolate between the EFA model (coordinated) and the classical random effect model (uncoordinated). By choosing the relative variances of the different components appropriately, we are able to simulate the standard additive model, the standard uncoordinated model, and the coordinated EFA model.

Finally, we use two alternative simulation frameworks for evaluating the calibration of our non-parametric sample splitting test with respect to LD and dominance that we described fully in the **Supplement**.

Yeast data

We downloaded genotype and phenotype data for prototrophic haploid segregants from a cross between a laboratory strain and a wine strain of yeast [13]. This dataset consisted of 46 quantitative traits for 1,008 samples genotyped at 11,623 unique markers. We standardized each phenotype and SNP to mean 0 and variance 1. To account for linkage disequilibrium in fixed effect models, we clumped+thresholded variants using threshold of $r^2 < 0.2$, minimum distance of 250 kb, and marginal additive $p < 0.05$ using PLINK 1.9 [77]. We performed clumping+thresholding separately on each cross-validated fold to avoid bias from overfitting [78].

Additionally, we downloaded a second yeast growth trait dataset from the same group with a larger sample size ($N = 4390$) [14]. This second dataset contained 20/46 of the growth phenotypes in the first dataset, and it included 28,220 unique genotype markers for each sample. We preprocessed this dataset in the same way as above.

We evaluated phenotype prediction accuracy with 10-fold cross validation, using the same 10 folds across all methods so we could test for differential prediction accuracy with paired t-tests.

UK Biobank data

We studied four traits in UK Biobank that were previously extensively characterized [44]: urate, IFG1, male testosterone, and female testosterone. The latter two traits are defined as testosterone levels measures in males or females. We analyzed white British individuals in the UK Biobank after removing related individuals identified by UKB. We studied the top 100 LD-pruned GWAS hits from [44], or all 77 GWAS hits for female testosterone. Prior to fitting EFA, we regressed out sex, age, batch, assessment center, and the top 10 genotype PCs.

To test statistical significance, we split the individuals into two halves and fit EFA on each. We then evaluate the squared correlation between estimated epistasis effect matrices (i.e., ω). We define an empirical null distribution for this metric by randomly permuting the residuals from the additive model within each half of the data, which we repeat 1000 times. It is important to regress out additive effects before permuting the phenotypes so that individuals are exchangeable. This is our first permutation test.

Second, to account for the arbitrary choice of seed, we repeat this procedure on 100 different seeds. For each seed, we evaluate one permutation of the samples, giving a comparable null distribution. We test for significant epistasis by asking if the real data have stronger epistasis than the permuted data across 100 seeds using a 1-sided, paired t-test. To assess robustness to phenotype scale, we repeat the entire process after first scaling the phenotype to match the quantiles of a normal distribution.

To test biological significance, we used the annotations of GWAS hits to biological pathways provided in [44]. We assigned SNPs to the epistasis factors by (1) polarizing SNPs to have positive additive effect sizes and (2) choosing the largest 15 effects on each pathway. To test robustness, we also evaluate results using 10 or 20 SNPs per pathway (**Supplementary Table 3-7**). We calculate p-values for enrichment of biological pathways in epistasis factors with a hypergeometric test.

References

[1] Cutting, G. R. Modifier genes in mendelian disorders: the example of cystic fibrosis. Ann. N. Y. Acad. Sci. **1214**, 57–69 (2010).

[2] Cutting, G. R. Cystic fibrosis genetics: from molecular understanding to clinical application. Nat. Rev. Genet. **16**, 45–56 (2015).

[3] Timberlake, A. T. et al. Two locus inheritance of non-syndromic midline craniosynostosis via rare SMAD6 and common BMP2 alleles. Elife **5** (2016).

[4] Starr, T. N. & Thornton, J. W. Epistasis in protein evolution. Protein Sci. **25**, 1204–1218 (2016).

[5] Bakerlee, C. W., Nguyen Ba, A. N., Shulgina, Y., Rojas Echenique, J. I. & Desai, M. M. Idiosyncratic epistasis leads to global fitness-correlated trends. Science **376**, 630–635 (2022).

[6] Corbett-Detig, R. B., Zhou, J., Clark, A. G., Hartl, D. L. & Ayroles, J. F. Genetic incompatibilities are widespread within species. Nature **504**, 135–137 (2013).

[7] Barton, N. H. How does epistasis influence the response to selection? Heredity (Edinb.) **118**, 96–109 (2017).

[8] Lin, X. et al. Nested epistasis enhancer networks for robust genome regulation. Science **377**, 1077–1085 (2022).

[9] Brem, R. B., Storey, J. D., Whittle, J. & Kruglyak, L. Genetic interactions between polymorphisms that affect gene expression in yeast. Nature **436**, 701–703 (2005).

[10] Shao, H. et al. Genetic architecture of complex traits: large phenotypic effects and pervasive epistasis. Proc. Natl. Acad. Sci. U. S. A. **105**, 19910–19914 (2008).

[11] Sittig, L. J. et al. Genetic background limits generalizability of genotype-phenotype relationships. Neuron **91**, 1253–1259 (2016).

[12] Huang, W. et al. Epistasis dominates the genetic architecture of drosophila quantitative traits. Proc. Natl. Acad. Sci. U. S. A. **109**, 15553–15559 (2012).

[13] Bloom, J. S., Ehrenreich, I. M., Loo, W. T., Lite, T.-L. V. & Kruglyak, L. Finding the sources of missing heritability in a yeast cross. Nature **494**, 234–237 (2013).

[14] Bloom, J. S. et al. Genetic interactions contribute less than additive effects to quantitative trait variation in yeast. Nature communications **6**, 8712 (2015).

[15] Mackay, T. F. C. Epistasis and quantitative traits: using model organisms to study gene–gene interactions. Nature Reviews Genetics **15**, 22–33 (2014).

[16] Forsberg, S. K. G., Bloom, J. S., Sadhu, M. J., Kruglyak, L. & Carlborg, Ö. Accounting for genetic interactions improves modeling of individual quantitative trait phenotypes in yeast. Nature Genetics **49**, 497–503 (2017).

[17] Jiang, Y. & Reif, J. C. Modeling epistasis in genomic selection. Genetics **201**, 759–768 (2015).

[18] Schrauf, M. F. et al. Phantom epistasis in genomic selection: On the predictive ability of epistatic models. G3 (Bethesda) **10**, 3137–3145 (2020).

[19] Hickey, K. L. et al. GIGYF2 and 4EHP inhibit translation initiation of defective messenger RNAs to assist ribosome-associated quality control. Mol. Cell **79**, 950–962.e6 (2020).

[20] Norman, T. M. et al. Exploring genetic interaction manifolds constructed from rich single-cell phenotypes. Science **365**, 786–793 (2019).

[21] Horlbeck, M. A. et al. Mapping the Genetic Landscape of Human Cells. Cell **174**, 953–967.e22 (2018).

[22] Dixit, A. et al. Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. Cell **167**, 1853–1866.e17 (2016).

[23] Carlborg, Ö. & Haley, C. S. Epistasis: too often neglected in complex trait studies? Nature Reviews Genetics **5**, 618–625 (2004).

[24] Hill, W. G., Goddard, M. E. & Visscher, P. M. Data and Theory Point to Mainly Additive Genetic Variance for Complex Traits. PLoS Genetics **4**, e1000008 (2008).

[25] Huang, W. & Mackay, T. F. C. The genetic architecture of quantitative traits cannot be inferred from variance component analysis. PLoS Genet. **12**, e1006421 (2016).

[26] Schrode, N. et al. Synergistic effects of common schizophrenia risk variants. Nat. Genet. **51**, 1475–1485 (2019).

[27] Weiner, D. J. et al. Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. Nature Genetics **49**, 978–985 (2017).

[28] Bergen, S. E. et al. Joint contributions of rare copy number variants and common SNPs to risk for schizophrenia. Am. J. Psychiatry **176**, 29–35 (2019).

[29] Manolio, T. A. et al. Finding the missing heritability of complex diseases. Nature **461**, 747–753 (2009).

[30] Hivert, V. et al. Estimation of non-additive genetic variance in human complex traits from a large sample of unrelated individuals. Am. J. Hum. Genet. **108**, 786–798 (2021).

[31] Cockerham, C. C. An Extension of the Concept of Partitioning Hereditary Variance for Analysis of Covariances among Relatives When Epistasis Is Present. Genetics **39**, 859 (1954).

[32] Henderson, C. R. Best linear unbiased prediction of nonadditive genetic merits in noninbred populations. Journal of Animal Science (1985).

[33] Sheppard, B. et al. A model and test for coordinated polygenic epistasis in complex traits. Proceedings of the National Academy of Sciences **118**, e1922305118 (2021). URL https://www.pnas.org/doi/abs/10.1073/pnas.1922305118. https://www.pnas.org/doi/pdf/10.1073/pnas.1922305118.

[34] Zuk, O., Hechter, E., Sunyaev, S. R. & Lander, E. S. The mystery of missing heritability: Genetic interactions create phantom heritability. Proceedings of the National Academy of Sciences of the United States of America **109**, 1193–1198 (2012).

[35] Young, A. I. et al. Estimating heritability without environmental bias. BioRxiv 218883 (2017).

[36] Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nature Genetics (2015).

[37] Finucane, H. et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. BioRxiv 103069 (2017).

[38] Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell **169**, 1177–1186 (2017).

[39] Liu, X., Li, Y. I. & Pritchard, J. K. Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. Cell **177**, 1022–1034.e6 (2019).

[40] Ma, Y., Patil, S., Zhou, X., Mukherjee, B. & Fritsche, L. G. ExPRSweb: An online repository with polygenic risk scores for common health-related exposures. Am. J. Hum. Genet. **109**, 1742–1760 (2022).

[41] Blair, D. R. et al. A nondegenerate code of deleterious variants in Mendelian loci contributes to complex disease risk. Cell **155**, 70–80 (2013).

[42] Vilhjálmsson, B. J. et al. Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. The American Journal of Human Genetics **97**, 576–592 (2015).

[43] Márquez-Luna, C. et al. Incorporating functional priors improves polygenic prediction accuracy in UK biobank and 23andme data sets. Nat. Commun. **12**, 6052 (2021).

[44] Sinnott-Armstrong, N., Naqvi, S., Rivas, M. & Pritchard, J. K. Gwas of three molecular traits highlights core genes and pathways alongside a highly polygenic background. eLife **10**, e58615 (2021). URL https://doi.org/10.7554/eLife.58615.

[45] Pazokitoroudi, A., Chiu, A. M., Burch, K. S., Pasaniuc, B. & Sankararaman, S. Quantifying the contribution of dominance deviation effects to complex trait variation in biobank-scale data. Am. J. Hum. Genet. **108**, 799–808 (2021).

[46] de Los Campos, G., Sorensen, D. A. & Toro, M. A. Imperfect linkage disequilibrium generates phantom epistasis (& perils of big data). G3 (Bethesda) **9**, 1429–1436 (2019).

[47] Hemani, G. et al. Detection and replication of epistasis influencing transcription in humans. Nature **508**, 249–253 (2014).

[48] Brown, A. A. et al. Genetic interactions affecting human gene expression identified by variance association mapping. eLife **3**, e01381 (2014).

[49] Fish, A. E., Capra, J. A. & Bush, W. S. Are interactions between cis-regulatory variants evidence for biological epistasis or statistical artifacts? Am. J. Hum. Genet. **99**, 817–830 (2016).

[50] Hemani, G. et al. Phantom epistasis between unlinked loci. Nature **596**, E1–E3 (2021).

[51] Sverdlov, S. & Thompson, E. A. The epistasis boundary: Linear vs. nonlinear genotype-phenotype relationships. bioRxiv (2018). URL https://www.biorxiv.org/content/early/2018/12/21/503466. https://www.biorxiv.org/content/early/2018/12/21/503466.full.pdf.

[52] Marchini, J., Donnelly, P. & Cardon, L. R. Genome-wide strategies for detecting multiple loci that influence complex diseases. Nature Genetics **37**, 413–417 (2005).

[53] Rawlik, K., Canela-Xandri, O. & Tenesa, A. Evidence for sex-specific genetic architectures across a spectrum of human complex traits. Genome Biol. **17**, 166 (2016).

[54] Khramtsova, E. A., Davis, L. K. & Stranger, B. E. The role of sex in the genomics of human complex traits. Nat. Rev. Genet. **20**, 173–190 (2019).

[55] Dahl, A. & Zaitlen, N. Genetic influences on disease subtypes. Annu. Rev. Genomics Hum. Genet. **21**, 413–435 (2020).

[56] Oliva, M. et al. The impact of sex on gene expression across human tissues. Science **369** (2020).

[57] Jannink, J.-L. Identifying Quantitative Trait Locus by Genetic Background Interactions in Association Studies. Genetics **176**, 553–561 (2007).

[58] Crawford, L., Zeng, P., Mukherjee, S. & Zhou, X. Detecting epistasis with the marginal epistasis test in genetic mapping studies of quantitative traits. PLoS Genetics **13**, e1006869 (2017).

[59] Akdemir, D. & Jannink, J.-L. Locally epistatic genomic relationship matrices for genomic association and prediction. Genetics **199**, 857–871 (2015).

[60] Rau, C. D. et al. Modeling epistasis in mice and yeast using the proportion of two or more distinct genetic backgrounds: Evidence for "polygenic epistasis". PLoS Genet. **16**, e1009165 (2020).

[61] Darnell, G., Smith, S. P., Udwin, D., Ramachandran, S. & Crawford, L. Partitioning tagged non-additive genetic effects in summary statistics provides evidence of pervasive epistasis in complex traits. bioRxiv (2022). URL `https://www.biorxiv.org/content/early/2022/09/11/2022.07.21.501001`. `https://www.biorxiv.org/content/early/2022/09/11/2022.07.21.501001.full.pdf`.

[62] Turchin, M. C., Darnell, G., Crawford, L. & Ramachandran, S. Pathway analysis within multiple human ancestries reveals novel signals for epistasis in complex traits. bioRxiv (2020). URL `https://www.biorxiv.org/content/early/2020/09/25/2020.09.24.312421`. `https://www.biorxiv.org/content/early/2020/09/25/2020.09.24.312421.full.pdf`.

[63] Demetci, P. et al. Multi-scale inference of genetic trait architecture using biologically annotated neural networks. PLOS Genetics **17**, 1–53 (2021). URL `https://doi.org/10.1371/journal.pgen.1009754`.

[64] Domingue, B. W., Kanopka, K., Mallard, T. T., Trejo, S. & Tucker-Drob, E. M. Distinguishing between interaction and dispersion effects in the analysis of gene-environment interaction. bioRxiv (2020). URL `https://www.biorxiv.org/content/early/2020/10/16/2020.09.08.287888`. `https://www.biorxiv.org/content/early/2020/10/16/2020.09.08.287888.full.pdf`.

[65] Domingue, B. W., Kanopka, K., Trejo, S., Rhemtulla, M. & Tucker-Drob, E. M. Ubiquitous bias and false discovery due to model misspecification in analysis of statistical interactions: The role of the outcome's distribution and metric properties. Psychol. Methods (2022).

[66] Ballard, J. L. & O'Connor, L. J. Shared components of heritability across genetically correlated traits. Am. J. Hum. Genet. **109**, 989–1006 (2022).

[67] Border, R. et al. Cross-trait assortative mating is widespread and inflates genetic correlation estimates. bioRxiv (2022). URL `https://www.biorxiv.org/content/early/2022/03/23/2022.03.21.485215`. `https://www.biorxiv.org/content/early/2022/03/23/2022.03.21.485215.full.pdf`.

[68] Udler, M. S. et al. Type 2 diabetes genetic loci informed by multi-trait associations point to disease mechanisms and subtypes: A soft clustering analysis. PLoS medicine **15**, e1002654 (2018).

[69] Liley, J., Todd, J. A. & Wallace, C. A method for identifying genetic heterogeneity within phenotypically defined disease subgroups. Nature Genetics **49**, 310–316 (2016).

[70] Werme, J., van der Sluis, S., Posthuma, D. & de Leeuw, C. A. An integrated framework for local genetic correlation analysis. Nat. Genet. **54**, 274–282 (2022).

[71] Kendler, K. S. & Gardner, C. O. Interpretation of interactions: guide for the perplexed. Br. J. Psychiatry **197**, 170–171 (2010).

[72] Gusev, A. et al. Quantifying Missing Heritability at Known GWAS Loci. PLoS Genetics **9**, e1003993–19 (2013).

[73] Gamazon, E. R. et al. A gene-based association method for mapping traits using reference transcriptome data. Nature Genetics **47**, 1091–1098 (2015).

[74] Mukamel, R. E. et al. Protein-coding repeat polymorphisms strongly shape diverse human phenotypes. Science **373**, 1499–1505 (2021).

[75] LaBianca, S. et al. Polygenic profiles define aspects of clinical heterogeneity in adhd. medRxiv (2021). URL `https://www.medrxiv.org/content/early/2021/07/15/2021.07.13.21260299`. `https://www.medrxiv.org/content/early/2021/07/15/2021.07.13.21260299.full.pdf`.

[76] Elliott, L. T. et al. Genome-wide association studies of brain imaging phenotypes in UK biobank. Nature **562**, 210–216 (2018).

[77] Purcell, S. et al. Plink: A tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics **81**, 559–575 (2007). URL `https://www.sciencedirect.com/science/article/pii/S0002929707613524`.

[78] Moscovich, A. & Rosset, S. On the cross-validation bias due to unsupervised preprocessing. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **84**, 1474–1502 (2022). URL `https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/rssb.12537`. `https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/rssb.12537`.