



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/220078/>

Version: Published Version

---

**Article:**

Balvert, M., Cooper-Knock, J., Stamp, J. et al. (2024) Considerations in the search for epistasis. *Genome Biology*, 25. 296. ISSN: 1474-7596

<https://doi.org/10.1186/s13059-024-03427-z>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

**Takedown**


If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.

REVIEW

Open Access



# Considerations in the search for epistasis

Marleen Balvert<sup>1\*†</sup>, Johnathan Cooper-Knock<sup>2\*†</sup> , Julian Stamp<sup>3</sup>, Ross P. Byrne<sup>4</sup>, Soufiane Mourragui<sup>5</sup>, Juami van Gils<sup>6</sup>, Stefania Benonisdottir<sup>7,19</sup>, Johannes Schlüter<sup>8</sup>, Kevin Kenna<sup>9</sup>, Sanne Abeln<sup>1,21</sup>, Alfredo Iacoangeli<sup>10,11,12</sup>, Joséphine T. Daub<sup>9</sup>, Brian L. Browning<sup>13</sup>, Gizem Taş<sup>1,9</sup>, Jiajing Hu<sup>10</sup>, Yan Wang<sup>9</sup>, Elham Alhathli<sup>2</sup>, Calum Harvey<sup>2</sup>, Luna Pianesi<sup>8</sup>, Sara C. Schulte<sup>14</sup>, Jorge González-Domínguez<sup>15</sup>, Erik Garrisson<sup>16</sup>, Lorentz workshop on epistasis, Michael P. Snyder<sup>17</sup>, Alexander Schönhuth<sup>8†</sup>, Letitia M. F. Sng<sup>18\*†</sup> and Natalie A. Twine<sup>18\*†</sup>

<sup>†</sup>Marleen Balvert and Johnathan Cooper-Knock contributed equally to this work.

<sup>†</sup>Alexander Schönhuth, Letitia M. F. Sng, and Natalie A. Twine contributed equally to this work.

\*Correspondence: m.balvert@tilburguniversity.edu; j.cooper-knock@sheffield.ac.uk; letitia.sng@csiro.au; Natalie.Twine@csiro.au

<sup>1</sup> Tilburg University, Tilburg, The Netherlands

<sup>2</sup> SITraN, University of Sheffield, Sheffield, UK

<sup>18</sup> Commonwealth Scientific and Industrial Research Organisation, Westmead, Australia

Full list of author information is available at the end of the article

## Abstract

Epistasis refers to changes in the effect on phenotype of a unit of genetic information, such as a single nucleotide polymorphism or a gene, dependent on the context of other genetic units. Such interactions are both biologically plausible and good candidates to explain observations which are not fully explained by an additive heritability model. However, the search for epistasis has so far largely failed to recover this missing heritability. We identify key challenges and propose that future works need to leverage idealized systems, known biology and even previously identified epistatic interactions, in order to guide the search for new interactions.

## Introduction

Epistasis refers to changes in the effect of a unit of genetic information (such as a single nucleotide polymorphism or a gene) on a phenotype, dependent on the context of other genetic units. Such interactions are biologically plausible and offer a potential explanation for phenomena not fully accounted for by an additive heritability model. Heritability is a measure of the extent to which phenotypic variation is genetically determined. Broad-sense heritability refers to heritability measured by comparison of concordance rates for phenotype between monozygotic and dizygotic twins who share 100% or 50% of their genetics, respectively [1]. Missing heritability commonly refers to the gap between measured broad-sense heritability and heritability calculated by adding together the individual contributions of phenotype-associated SNPs genome-wide (i.e., narrow-sense heritability). Missing heritability is important because it implies that we have an incomplete understanding of the genetic basis of health and disease. A number of possibilities could explain this missing heritability, including gene-environment interactions. Epistatic interactions are another candidate to explain a proportion of missing heritability but an alternative explanation is



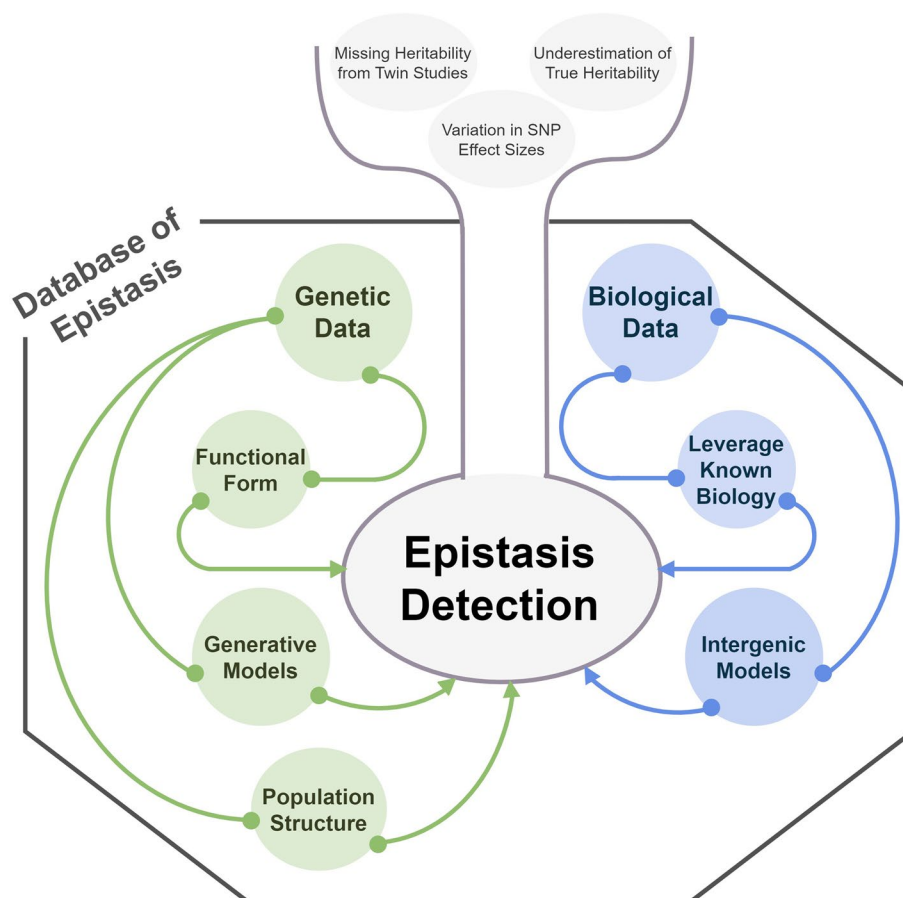
© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

that current knowledge is simply missing the statistical power to discover all important additive effects. However, there is good observational evidence for epistasis, for example, from large-scale screens in yeast studying the effect of combinations of individual gene knockouts [2, 3].

A meta-analysis of twin studies concluded that for 69% of traits the data was consistent with an additive model whereby monozygotic twin correlations were almost exactly double dizygotic twin correlations [4]. However, even this study provides evidence for non-additive genetic effects in a subset of traits. For traits such as depressive disorder, hyperkinetic disorders, and atopic dermatitis, the authors observed monozygotic twin correlations which were greater than double the dizygotic twin correlations, consistent with a non-additive genetic effect. Moreover, even observations consistent with an additive model are not equivalent to actually demonstrating an additive model and the presence of an additive model does not necessarily rule out the possibility of an underlying epistatic model. Interestingly, the effect sizes of a majority of SNPs vary between genetic backgrounds [5], suggesting the presence of interactions between the genetic background and the SNP. Finally, in simulations of epistasis, additive models used to measure narrow-sense heritability fail to account for non-linear interactions between genetic variants and thus dramatically underestimate true heritability [6].

The problem is that previous searches for epistasis have so far largely failed to recover missing heritability [7]. Various computational approaches using statistics, combinatorics, and machine learning have been applied to try and detect epistasis. Each of these approaches try to address the issue of identifying relevant potential epistatic interactions from an enormous search space, either by enumerating all possibilities or by finding an efficient way to move through the search space. Consideration of epistasis inherently leads to a combinatorial explosion: the number of potential interactions increases exponentially with the number of genetic characteristics involved in each interaction.

During a workshop entitled “*A multidisciplinary approach to epistasis detection*,” held at the Lorentz Center in The Netherlands in July 2023, 41 experts on epistasis detection from a variety of fields came together. Through interactions and discussions, we identified challenges that need to be addressed in order to advance epistasis detection. We consider the central combinatorial challenge of epistasis identification through two perspectives: statistical and mathematical approaches to case–control studies versus leveraging biological knowledge and models (Fig. 1). Each of the two perspectives is addressed through three subtopics. For the statistical and mathematical perspective, we start by reviewing specific problems with popular model assumptions and pose the question of whether it is possible to avoid assuming any mathematical form. Next, we discuss the potential of novel generative AI models for the analysis of case–control cohort data. Third, we show empirically the importance of accounting for population structure in case–control cohort studies, which unfortunately is often overlooked. In the second half of this review, we discuss biological observations of epistasis. We start with the idea that search for epistasis should always start with biological models. Second, we discuss whether one should consider inter- and intragenic epistasis separately. Finally, we propose the use-case for a “database of epistasis” and provide guidelines for the characteristics that such a database should have.



**Fig. 1** Key considerations for a comprehensive consideration of epistasis. Epistasis is poised to play a key role in genetic architecture and in the missing heritability problem. In this review, we look at epistasis from two perspectives: driven by genetic information (green circles) and by biological observations (blue circles). We discuss how genetic data can be used in a functional form or in generative models to detect epistasis, and that the inclusion of population structure information derived from genetic data is crucial. On the other hand, the discovery of epistasis can also be informed by biological observations. Ideally, both sides will lead to better detection of epistasis, ultimately leading to a key resource that is a database of epistatic interactions

### What assumptions of epistasis are being made and what are their implications?

Epistasis is a natural expectation of a complex system, but the search for epistasis is challenging primarily due to the combinatorial explosion of possibilities. In this section, we delve into the assumptions, mathematical or otherwise, that underpin current methods of epistasis detection and posit the use of state-of-the-art machine learning approaches in a new generation of data-driven epistasis detection methods. Many existing approaches have been recently reviewed [8]; here we extend this analysis by considering the conceptual limitations of current works and more novel approaches.

#### Generalizing the functional form of epistasis

If epistasis is taken in its statistical sense as the deviation from the additive/linear baseline, then all other terms—namely quadratic and higher order interactions—are epistasis [9]. The relation between genotype and phenotype can then be represented as a function

that maps a discrete sequence space onto one or more binary- or real-valued traits. Extending the formulation used in [10], a phenotype impacted by epistasis can be formulated mathematically as:

$$y = \sum_{a \in A} \beta_{\alpha(a)} \prod_{i \in (1, \dots, N)} x_i^{a_i}, \quad (1)$$

where  $N$  is the total number of SNPs in the data,  $x_i$  encode the SNP information (e.g., allelic dosage),  $y$  symbolizes the phenotype, and

$$A := \{a \in \{0,1\}^N : 1^T a \leq d\}$$

with  $d$  the order of the highest-order interaction. The parameter  $d$  allows one to choose a maximum order for the epistatic interaction, which can be at most  $N$ .  $\beta_{\alpha(a)}$  are the parameters to be estimated denoting the magnitude of the epistatic effect of the variants corresponding to the vector  $a$ , where  $\alpha(a)$  is the index corresponding to the vector  $a$  if one were to order all elements of  $A$ . The vector  $a$  thus indicates which variants are included in the  $\alpha(a)$  th interaction. For example, in the case of  $N = 3$  and  $d = 2$  this would give:

$$\begin{aligned} y &= \beta_0 x_1^0 x_2^0 x_3^0 + \beta_1 x_1^1 x_2^0 x_3^0 + \beta_2 x_1^0 x_2^1 x_3^0 + \beta_3 x_1^0 x_2^0 x_3^1 + \beta_4 x_1^1 x_2^1 x_3^0 + \beta_5 x_1^1 x_2^0 x_3^1 + \beta_6 x_1^0 x_2^1 x_3^1 \\ &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1 x_3 + \beta_6 x_2 x_3. \end{aligned} \quad (2)$$

Note that since  $d = 2$ , the interaction between all three variants is not included.

In other words, epistasis is the combined effect of any combination of SNPs up to a certain order of magnitude. For binary traits, one can apply the logit function to the right-hand side of (1). Note that explicitly using formulation (1) leads to a combinatorial explosion in the number of terms and hence parameters to be estimated as the number of SNPs and the degree  $d$  increase, hence explicitly estimating the effect of all interaction terms is infeasible.

Any function can be represented as a series expansion with the commonly used Taylor and Fourier series expansion [11]. The difference between the two representations for modeling epistasis is the reference frame. The Taylor series uses the wild type as reference to quantify epistatic interactions and in the Fourier series epistatic effects are averages over all backgrounds [12]. With epistasis, a wide body of literature suggests that many different mathematical formulations can be linked using the weighted Walsh-Hadamard transform [12].

Models that identify epistatic interactions from genotype data nearly always make assumptions on the form of the epistatic relationship (Table 1). There is a gradient in current approaches of epistasis detection: from models assuming a specific form of epistasis (e.g., BOOST [13], BitEpi [14], Fiucho [15], IRELAND [16, 17], MDR [17]) to models that learn an epistatic relationship of any form (e.g., [12, 18–20]). These assumptions are described in the paragraphs below.

Approaches that assume a specific form of epistatic interaction, for example, pairwise, triplet, or quadruplet interactions, are often easier to understand and can provide directly interpretable outcomes. However, they still suffer from the combinatorial explosion if there is no constraint on the type and number of interacting variants.

**Table 1** Table summarizing the assumptions made by various methodological approaches and tools for detecting epistasis

Assumption	Approach/tool
Two-way interaction only	BOOST, GenEpi, EpiGWAS, MDR
Two-, three-, or four-way interactions	BitEpi (4-way), LOBICO (4-way), MDSN (3-way)
Any order of interaction	IRELAND, Fiuncho
Highly polygenic/omnigenic trait	Deep neural nets
SNP interaction partners are independent	Random forests; any tool which does not account for population structure
Variant interactions can be described through a combination of decision trees	Random forest
Cases can be separated from controls through a hyperplane in the (modified) variant space	Support vector machine
Epistatic interactions are described through logic statements	LOBICO, IRELAND

Consequently, several methods focus on two-way interactions only [13, 21, 22] while other more exhaustive search methods are limited by the computational complexity of the approach and often do not go beyond four-way interactions [14, 15, 23, 24]. Fiuncho and IRELAND do go beyond four-way interactions [14–17, 23, 24], though they are limited in the number of SNPs they can analyze simultaneously. Whether going beyond four-way interactions is clinically relevant and can be validated beyond statistical evidence remains an open question. It is, however, biologically plausible that many SNPs are involved in the same epistatic interaction [25, 26].

On the other hand, freeform approaches such as deep neural networks (DNNs) [19, 20, 27, 28], as supported by mathematical theorems (the universal approximation theorem, e.g., [29, 30]), can approximate arbitrary functional relationships, thereby in theory they can avoid the requirement to impose any assumption in terms of functional relationships driving epistasis. This makes them more flexible and less prone to computational limits. In practice, however, these approaches require additional steps, perhaps yet to be developed, to not only implicitly capture but also provide an explicit description of the epistatic interaction [18]. Because DNNs tend to use a large number of input variables for phenotype prediction, they arguably assume a highly polygenic or even omnigenic trait [29] in practice.

Many classical machine learning (ML) approaches sit between these two extremes, such as decision tree ensembles, i.e., random forest [31–35], boosting [36], and support vector machines [37–40]. These approaches make mild assumptions on the functional form of the epistatic interaction, allowing them to deal with higher-order interactions. For random forests, these assumptions include that SNPs forming interactions must be independent of each other. Random forests also assume that the relationship between epistasis and genetic variants can be described as a combination of decision trees, while support vector machines assume that one can separate cases from controls by using a hyperplane in the (transformed) variant space.

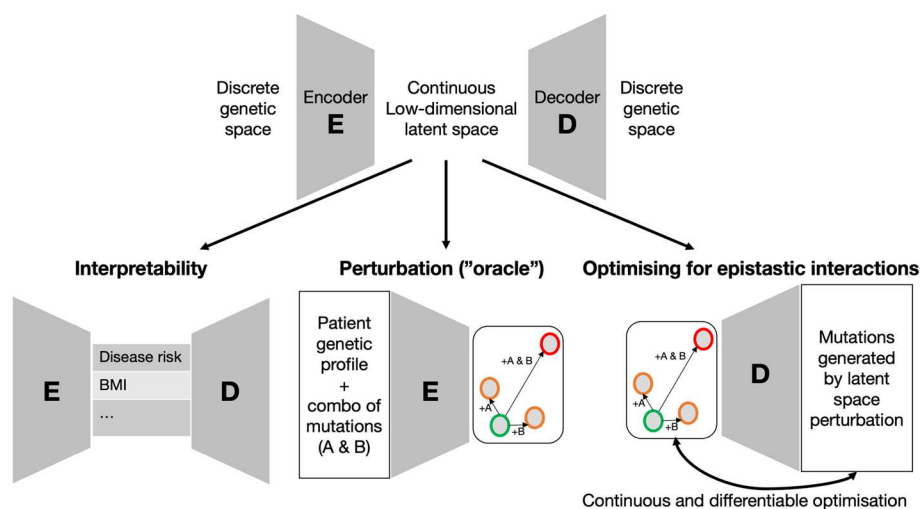
The standard formulations of epistasis presented above link genotype to phenotype by means of statistical models, machine learning approaches, or combinatorics, all based on large datasets. However, the joint probability structure of the dataset is never explicitly exploited during inference. Modelization and subsequent dissection of the joint

probability between genotypic features, and between genotypic features and a phenotype of interest, offer another approach to study epistasis.

### Using generative approaches to model and explore epistasis

Apart from the classical ML approaches aforementioned, recent algorithmic and computational advances have offered insight into the potential of deep generative models in genetics [41, 42]. Inspired by many successful applications in other scientific fields [43–45], we envision that leveraging generative approaches could offer a transformative approach to identify genetic interactions. Note that, just as for DNN based classifiers, universal approximation theorems for probability distributions [46] support the utmost flexibility of generative deep learning approaches. A large part of the advantage of this approach rests on the ability to perturb a latent space representation of genetic interactions and make observations regarding the effect on phenotypes: in effect providing an experimental system with a tractable number of variables. We explain and explore this idea below.

A deep generative model aims to construct a condensed representation of the genetic information that accurately describes the distribution of genetic variance in the population from which observed genetic data is sampled. That is, the generative model learns how to represent an individual's genetic information in a condensed manner with minimal loss of information, meaning that it can reconstruct the original genetic information from this condensed representation with high accuracy. Deep generative models are typically composed of three components: the encoder, the latent space, and the decoder



**Fig. 2** Example of how generative modeling can be employed to hunt for genetic interactions. Most deep generative models are made of two elements: the encoder, which reduces dimensionality, and the decoder, which can generate genetic profiles *in silico* (top panel). We present three problems where generative models can be employed. Interpretability: The output of the encoder, and input of the decoder, can be interpreted and related to phenotypes of interest. Perturbation: A patient's genetic profile can be perturbed *in silico* and passed through the encoder. For instance, a patient with two wild-type alleles (green circles) can be modified by induction of A or B (orange circles) or both at the same time (red circles). Study of the corresponding perturbation in the latent space can help prioritize potentially interacting genetic pairs. Optimization: Finally, a deep generative model could be directly employed inside an optimization strategy geared towards finding epistatic interactions, benefiting from two advantages of deep generative models: the auto-differentiation of the decoder and the continuous character of the latent space

(Fig. 2). Firstly, the encoder maps a sample to a space of much lower dimensionality. This so-called latent space offers an intriguing property—it is continuous, unlike the binary nature of genetic profiles (wild-type or mutated). Finally, the decoder takes a point within this latent space, whether it originates from the encoder or is chosen randomly, and reconstructs the corresponding sample. Note that this sample could be non-existent if the point in the latent space is chosen randomly. Although deep generative models come in various styles [47–49], the roles of the encoder and decoder may differ, but they all share these fundamental characteristics.

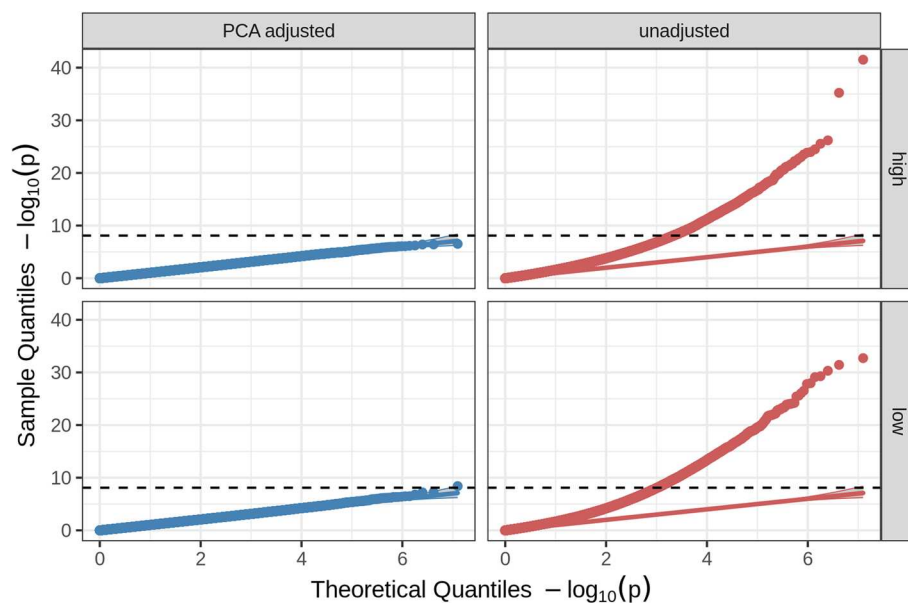
We foresee three kinds of applications that require the development of models with a disentangled latent space [48, 50]. Firstly, extrapolating from protein structure work, we expect that well-designed models will exhibit emergent information in the latent space [51]. In more concrete detail, using interpretable labels (such as diseased/non-diseased), one trains the encoder to map differently labeled data to distant parts of the latent space, while placing identically labeled data near to each other. Such training procedures can be implemented by the integration of contrastive loss functions [52] into the training of the encoder-decoder architecture. As per their definition, contrastive loss functions are exactly the drivers that keep similar things together, while keeping separate things apart when embedding data into latent space. As a result, the interpretability of the model obtained by an appropriately structured continuously valued latent space could be instrumental in increasing the power of standard analysis (Fig. 2, Interpretability). For example, linking parts of the latent space to known phenotypes (e.g. disease risk) could aid in identifying new disease-risk regions. This expansion of the available dataset would enhance the power of standard epistasis detection analyses. A second, more direct application of such a model involves using it as an “oracle” that provides quantitative insights into the perturbation caused by a pair of genetic alterations (Fig. 2, Perturbation). For instance, given two alleles A and B at different loci, one could measure the perturbation in the latent space induced by each mutation and compare it to the perturbation in the latent space caused by A and B combined. If the combined mutation leads to the same perturbation as the two individual combinations together, then there is no indication of epistasis, else there is. Lastly, the model can be used in a more exploratory manner through the design of optimization routines (Fig. 2, Optimization). Using the decoder’s gradients enables the identification of genetic pairs that lead to maximal perturbations in the latent space, indicating interaction within these pairs and hence identifying potential epistatic interactions. These three directions, far from being exhaustive, showcase the potential of deep generative models in the detection of epistatic interactions. Since the use of these methods in genomic applications is still in its infancy but highly promising, extensive further research along these lines is necessary.

### **Population structure confounds regression-based epistasis detection**

In addition to the assumptions on the form of the epistatic interaction, there are underlying assumptions of genetic data that should inform epistasis detection models, particularly linkage disequilibrium (LD). However, many epistasis detection datasets and tools fail to account for LD structures which means they will be particularly vulnerable to population mismatch. Here we include a detailed consideration of this failure within epistasis detection and how this could be addressed.

Events in human evolutionary history such as migration and admixture [53] are reflected in differences in allele frequencies (AFs) between different populations [54]. The concept of genetic populations is a simplified description of these genetic patterns [55]. The differences in AFs between populations are called population structure and have been described as a confounding factor in genome-wide association studies (GWAS) [56–59]. In a naive association test, the samples are modeled as independent, an assumption that cannot hold when there are such systematic genetic trends within the data. Epistasis analysis, similar to GWAS, is vulnerable to confounding from population structure, which if uncorrected, can result in substantial  $p$  value inflation and false positives in analyses with no true epistatic interactions (Fig. 3). Furthermore, previous research has shown that a slight change in the AF of a SNP results in a substantial decrease in power to replicate the main effects of said SNP when there is an underlying epistatic model [60]. Detection and correction of population structure are thus of core importance to the study of epistasis. We propose that solutions to this problem can be informed by common practices from GWAS analysis.

In GWAS, there are two main approaches to correcting for population structure. The first includes principal components of genetic similarity as additional covariates in a linear model. Principal component analysis (PCA) aims to explain the variance–covariance structure of a high-dimensional data set with a relatively small number of linear



**Fig. 3** Population structure confounds regression-based epistasis detection. QQ plots for PLINK pairwise epistasis analysis on simulated null data with population structure and no true epistatic effects (interactions), i.e., only additive contributions (main effects), and trait heritability of 0.5 (details of simulation in Additional file 1: Supplementary Methods). Comparison of analysis corrected for population structure with 20 PCs (blue) and uncorrected for population structure (red) shows that population structure also leads to inflation of small  $p$  values and a large number of false positives in regression-based genome-wide association studies that model epistasis as a pairwise interaction term. The dashed horizontal line is the significance threshold after Bonferroni correction. Phenotype adjustment here is performed by regressing the phenotype against 20 PCs in a multiple linear regression model and using the residuals as the “adjusted phenotype.” The facets labeled as “high” and “low” correspond to 1000 and 100 true causal variants respectively with additive-only contributions to trait variance

combinations of the original variables [61]. The first few principal components of genetic data often capture population structure and are suitable covariates for correcting this source of confounding [62]. The second includes a random effect that is informed by the genetic covariance between samples in a linear mixed model (LMM) approach. By including a random effect that covaries with the genetic similarity, the samples are no longer modeled as independent. This method, while computationally more costly, is able to account for population structure without overfitting; in the presence of cryptic relatedness LMMs outperform principal component-based correction methods [57, 63].

Analogous methods to those used in GWAS have been adopted to account for population structure in some epistasis detection approaches, including methods adopting PCA correction (e.g., MBMDR-PC [64]) and LMMs (e.g., REMMA-epi [65] and FaST-LMM-epi [66, 67]). Indeed, LMM approaches have been shown to produce significantly lower statistical inflation than PLINK's pairwise epistasis method [65, 68]. Surprisingly, several commonly used methods for epistasis detection including PLINK epistasis and BOOST [13] do not have a built-in option for including covariates or otherwise correcting for population structure.

Our simulation analysis suggests that simply ignoring population structure in these cases is unwise and would lead to substantial statistical inflation and false positives (Fig. 3, Additional file 1: Supplementary Methods). Here we simulated traits with no true epistatic effects (only additive effects) in a structured population and performed plink epistasis detection to evaluate the impact of population structure on the resulting test statistics. We expect that if epistasis tests were inherently robust to confounding from population structure there would be no significant hits or  $p$  value inflation, as no epistasis was simulated. For population structure correction, the phenotype was adjusted using multiple linear regression on the first 20 PCs prior to analysis. The simulation code is open source at [https://github.com/jdstamp/leiden\\_paper](https://github.com/jdstamp/leiden_paper). QQ plots of our simulations show evidence of statistical inflation and large numbers of false positives only in the simulations with no correction for population structure (Fig. 3, “unadjusted” panels on right), indicating that population structure can confound epistasis detection methods, while analyses corrected for structure were well-controlled (Fig. 3, “PCA adjusted” panels on left). We thus recommend that researchers using methods without built-in population structure correction for epistasis analysis address population structure, for example, by first adjusting their phenotype using principal components (taking the residuals from a multiple regression), or alternatively using a suitable LMM approach.

### **Leveraging biology in the search for epistasis**

The first part of this review focused on using statistical and mathematical approaches to identify epistasis from data. In this second part, we focus on if and how biological information can be leveraged to look for epistasis. We pose two questions surrounding the use of model systems and intergenic versus intragenic mechanisms in the search for epistasis, followed by a discussion on the usefulness of a database of epistasis.

#### **Should a search for epistasis start with biological observations?**

We assert that conclusive evidence for the role of epistasis in determining disease heritability has come from interactions that have been identified in large case–control cohorts

and model systems such as cell lines and organisms. We hypothesize that true epistatic interactions will be observable in both model systems and case–control cohorts. However, false positive epistatic interactions may be more likely in case–control datasets where, for example, population structure is imperfectly matched. On the other hand, the potential challenge with model systems is knowing whether the readout and the cell/tissue context are a correct approximation of disease. Indeed, a model system may suggest an epistatic interaction which is specific to the genetic background of the model organism and may not be important in an outbred population.

Many double mutant genetic knockout screens have been performed, both in model organisms and human cell lines [69]. The problem with the use of organisms for epistasis detection is, again, the size of the search space. Extensive large-scale screens have been performed in yeast, focusing on large-scale characterization of cells with combinations of two or more individual gene knockouts or temperature alleles [2, 70, 71]. Such screens have proved useful, for example, in delineating biological pathways containing genes with similar interaction profiles. However, they have not, as yet, provided the scale necessary for an exhaustive search for epistasis. Indeed higher organisms such as mice are not at all suitable for large-scale screens due to the practical undertaking involved in exhaustive characterization.

Unlike organisms, cells are more tractable. In mammalian cells, many double mutant screens have been performed, mostly in the context of cancer where gene knockouts have important therapeutic implications since gene knockouts can represent drug-targeting conditions. For example, in one recent study double mutant screens were performed for ~34,937 gene pairs in MCF-10A breast cell lines, and their effects on tumor growth were examined in mice [72]. Statistically significant gene pairs were identified and grouped into interacting modules. Interestingly, the genes within a group exhibited epistatic effects on gene expression of other group members. Overall, this study revealed the gene interaction network of tumor growth and has important implications for therapeutic strategies. Another recent work examined 1191 putative functional gene pairs and/or paralogs in human melanoma lines and identified 109 pairs that affected fitness [73]. An important consideration in a biological experiment is context, for example, epistatic interactions may only be apparent in a specific environment; when that environment is the presence of a particular toxin or therapeutic, this observation can be used to identify epistatic interactions which have the potential to guide personalized medicine [74].

An important and well understood biological consideration is the separation between intergenic and intragenic mechanisms of epistasis. There is good evidence for intragenic interactions such as haplotypes associating with altered gene expression depleted for deleterious coding alleles [75]. Similarly, there is evidence for intergenic epistasis, particularly between genes of similar function [76], where an established example concerns mutations within different hemoglobin beta-chains [77]. Intergenic epistasis across different functional pathways also tends to be the result of compensatory adaptation [78] and typically genes within the same pathway have a similar interaction profile [70]. The problem with separating intergenic and intragenic epistasis is that both are combined in real-world biological systems. To circumvent this, we suggest a stepwise approach where

intergenic epistasis is analyzed and identified before searching for intragenic epistasis within each intergenic interaction.

An alternative framing is that the search for epistasis should prioritise variants where biological evidence for an effect is provided by a case-control cohort instead of a model system. If there is an epistatic interaction, we might expect to be able to measure the association between phenotype and genotype even with only one of the involved variants. Intuitively, this will depend on the frequency of the alleles in question, the size of the study population, and the effect size of the genetic variants. If true, then we should be able to use prioritization methods based on independent models (i.e. an additive model) to reduce the search space size for epistasis. One study has already applied this principle [21] where the search for epistasis was focused on additional genetic variants which increased the effect of another, significant in isolation, genetic variant. Their results are promising, showing that in several datasets this method outperforms GBOOST and Lasso. A future extension for this approach might be to use symbolic regression [79, 80] to detect epistasis between genetic variants with nominal significance and to determine the mathematical formulation of the relationship without a pre-specification.

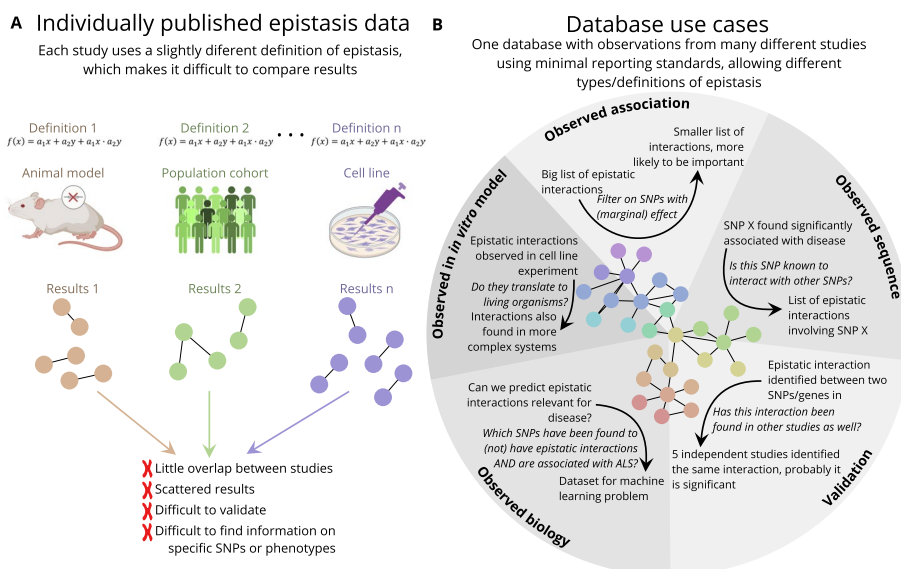
#### **A database of epistasis**

An obvious approach to leveraging known information in the search for epistasis is to use the literature of known epistatic interactions. There have been efforts to collect large amounts of epistasis data in one platform. For example, using pre-selected gene-specific transcription factors in *Saccharomyces cerevisiae* [3], SynLetDB provides a database specifically for synthetic lethality cases. However, most studies report genome-wide epistasis which are restricted to one organism or are phenotype-specific such as Alzheimer's disease [81, 82] or in cancer [83], and are not collated and standardized into a single database. A database that does exist for epistasis across multiple diseases is driven by a single methodological approach (<https://epistasis-disease-atlas.com>, [84]), therefore is limited in the types of epistatic interactions it contains. Furthermore, it remains difficult for researchers to reuse epistasis data because of the different definitions of epistasis, and different experimental and computational techniques used to identify epistasis.

We argue that although studies on epistasis can be highly diverse, there is a core set of data and metadata that can and should be reported for all studies to be able to effectively leverage known epistatic interactions. This set of minimum reporting standards could be based on other guidelines available such as MIAME/MINSEQE [85] that are used in sequence-based platforms, for example, the EGA (<https://web2.ega-archive.org/>) or GEO ([www.ncbi.nlm.nih.gov/geo/](http://www.ncbi.nlm.nih.gov/geo/)), and should include metadata per study, per sample, and per interaction, as outlined in Table 2. If all studies publishing epistasis information adhere to the same set of minimal reporting standards, referencing and using known epistasis would become much simpler. Additionally, it would facilitate the collection of epistasis information into one large database, which would benefit many researchers in the field of epistasis. Because epistasis is such a complex phenomenon, generating a database would be helpful to explore available data or validate new results (Fig. 4B). Additionally, a database could identify genes of interest, or other studies using a specific method of epistasis detection.

**Table 2** The set of minimum reporting standards for a database of epistasis

Per study	[1] Phenotype description [2] Experimental method [3] Computational method [4] Study sample size
Per sample	[1] Organism/cell line [2] Sample description
Per interaction	[1] SNP ID or SV gene loci [2] Associated HUGO gene symbol [3] Value of phenotype measure (e.g., gene expression, disease, trait, cell growth) [4] Confidence or significance score [5] Validation status [6] Nature of interaction (e.g., pairwise/higher order)



**Fig. 4** The need for an epistasis database. **A** It is currently difficult to compare epistasis data between studies, as there are many different approaches, models, and even definitions of epistasis. **B** Use cases of an epistasis database. By collecting epistasis data into one coherent framework, researchers can more easily find relevant information about their genes/SNPs and interactions of interest. Additionally, collecting epistasis data into one large framework would benefit from the creation of a reporting standard for epistasis data, such that epistasis data can be more easily collected and reused in the future

It is important to consider which types of researchers may use a database of epistasis. For example, researchers studying a certain gene or SNP may want to query that gene/SNP to get an overview of potential epistatic interactions, while studies identifying epistatic interactions may query specific interactions to validate their findings, or search for orthogonal sources of validation, e.g., knock-out studies for specific interactions demonstrating a phenotypic effect or reduced expression level at the mRNA or protein level. On the other hand, researchers focusing on a specific disease or other phenotype may query that phenotype to find any associated interactions.

However, because epistasis is such a diverse phenomenon, creating a comprehensive database to capture this information poses several challenges. For example, it would be ideal for a database to contain curated positive (i.e., epistatic interaction occurs) and negative (i.e., epistatic interaction does not occur) cases. A further complication is that computational methods use a range of measures of confidence for epistatic interactions

(e.g., statistical significance ( $p$  value/FDR), feature weight, knock-out experiments), and thus defining and standardizing positive and negative epistatic interaction is not straightforward. Likewise, these computational methods use different approaches (e.g., neural nets, random forests, regression), and thus have different metrics to detect epistasis, which may not be directly comparable (Fig. 4A). Hence, the provenance of each epistatic interaction, as suggested in Table 2, is essential to filter data and find epistatic interactions with multiple sources of evidence.

## Conclusion

The search for epistatic interactions which influence disease heritability is challenging but essential to facilitate effective personalized medicine. We have outlined some of the challenges, particularly the intractability of modeling epistasis. While we have suggested some modeling approaches that may be fruitful in the future, we have also considered real steps that could be used to improve models by integrating known biology. In particular, we suggest that all models of epistasis in the future should include an explicit correction for population structure and we show the importance of this through simulation. We make the case for a database of epistasis to bring together what is already known in the hope that this is a firmer foundation for discovery than strict model definitions, which may or may not be representative. Using generative models may be a way in which these biological observations can be summarized in a meaningful fashion.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-024-03427-z>.

Additional file 1. Supplementary methods for simulation of the effect of population structure on detection of epistasis, based on models in [88] and [89].

Additional file 2. Review history.

## Acknowledgements

This work is the result of fruitful discussions that took place at the workshop “A multidisciplinary approach to epistasis detection” in July 2023, in Leiden, the Netherlands. We thank the Lorentz Center for their support in facilitating and co-funding the workshop.

Additional members of the Lorentz workshop on epistasis

Name	Affiliation
Ammar Al-Chalabi	King's College London
Jorge Avila Cartes	Università degli Studi di Milano-Bicocca
Jasmijn Baaijens	Delft University of Technology
Joanna von Berg	Princess Maxima Center for Pediatric Oncology
Davide Bolognini	Fondazione Human Technopole
Paola Bonizzoni	Università degli Studi di Milano-Bicocca
Andrea Guarracino	University of Tennessee
Mehmet Koyuturk	Case Western Reserve University
Magda Markowska	University of Warsaw
Johannes Schlüter	Bielefeld University
Raghuram Dandinasivara	Bielefeld University
Jasper van Bemmelen	Delft University of Technology
Sebastian Vorbrugg	Max Planck Institute for Biology
Sai Zhang	University of Florida
Bogdan Pasanuic	University of Pennsylvania

**Peer review information**

Andrew Cosgrove was the primary editor of this article at *Genome Biology* and managed its editorial process and peer review in collaboration with the rest of the editorial team.

**Review history**

The review history is available as Additional file 2.

**Authors' contributions**

All authors contributed to the writing of the manuscript. All authors read and approved the final version.

**Funding**

We also thank the Artificial Intelligence Journal, the Company of Biologists, the Netherlands Organization for Scientific Research (Veni grant VI.Veni.192.043), Bielefeld University, Ronin, and funding from the European Union's Horizon 2020 research and innovation programme under Marie Skłodowska-Curie grant agreements No 956229 (ALPACA) and No 872539 (PANGAIA) for providing funding for the workshop. R.P.B. is supported by funding from the Motor Neurone Disease Association (Byrne/Oct22/979–799). J.C.K. was supported by the Wellcome Trust (216596/Z/19/Z). C.H./J.C.K. are supported by the MND (899–792). M.B. is supported by the Netherlands Organization for Scientific Research (Veni grant VI.Veni.192.043). M.P.S. is supported by the National Institutes of Health (CEGS 5P50HG00773504, 1P50HL083800, 1R01HL101388, 1R01-HL122939, S10OD025212, P30DK116074, and UM1HG009442).

**Data availability**

The code for the simulated data is available at Github [86] and Zenodo [87] under an MIT license.

**Declarations****Competing interests**

The authors declare that they have no competing interests.

**Author details**

<sup>1</sup>Tilburg University, Tilburg, The Netherlands. <sup>2</sup>STraN, University of Sheffield, Sheffield, UK. <sup>3</sup>Brown University, Providence, USA. <sup>4</sup>Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Ireland. <sup>5</sup>Hubrecht Institute, Utrecht, The Netherlands. <sup>6</sup>Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. <sup>7</sup>University of Oxford, Oxford, UK. <sup>8</sup>Bielefeld University, Bielefeld, Germany. <sup>9</sup>UMC Utrecht, Utrecht, The Netherlands. <sup>10</sup>Department of Biostatistics and Health Informatics, King's College London, London, UK. <sup>11</sup>Department of Basic and Clinical Neuroscience, King's College London, London, UK. <sup>12</sup>NIHR BRC SLAM NHS Foundation Trust, London, UK. <sup>13</sup>University of Washington, Seattle, USA. <sup>14</sup>Algorithmic Bioinformatics and Center for Digital Medicine, Heinrich Heine University, Düsseldorf, Germany. <sup>15</sup>CITIC, University of A Coruña, A Coruña, Spain. <sup>16</sup>University of Tennessee, Knoxville, USA. <sup>17</sup>Department of Genetics, Stanford University, Stanford, USA. <sup>18</sup>Commonwealth Scientific and Industrial Research Organisation, Westmead, Australia. <sup>19</sup>University of Iceland, Reykjavik, Iceland. <sup>20</sup>Department of Epidemiology, University of Florida, Gainesville, FL, USA. <sup>21</sup>Utrecht University, Utrecht, The Netherlands.

Received: 1 March 2024 Accepted: 23 October 2024

Published online: 19 November 2024

**References**

1. Verweij KJH, Yang J, Lahti J, Veijola J, Hintsanen M, Pulkki-Råback L, et al. Maintenance of genetic variation in human personality: testing evolutionary models by estimating heritability due to common causal variants and investigating the effect of distant inbreeding. *Evolution*. 2012;66:3238–51.
2. Segrè D, Deluna A, Church GM, Kishony R. Modular epistasis in yeast metabolism. *Nat Genet*. 2005;37:77–83.
3. Sameith K, Amini S, Groot Koerkamp MJA, van Leenen D, Brok M, Brabers N, et al. A high-resolution gene expression atlas of epistasis between gene-specific transcription factors exposes potential mechanisms for genetic interactions. *BMC Biol*. 2015;13:112.
4. Polderman TJC, Benyamin B, de Leeuw CA, Sullivan PF, van Bochoven A, Visscher PM, et al. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet*. 2015;47:702–9.
5. Patel RA, Musharoff SA, Spence JP, Pimentel H, Tcheandjieu C, Mostafavi H, et al. Genetic interactions drive heterogeneity in causal variant effect sizes for gene expression and complex traits. *Am J Hum Genet*. 2022;109:1286–97.
6. Li J, Li X, Zhang S, Snyder M. Gene-environment interaction in the era of precision medicine. *Cell*. 2019;177:38–44.
7. Mackay TF, Moore JH. Why epistasis is important for tackling complex human disease genetics. *Genome Med*. 2014;6:124.
8. Russ D, Williams JA, Cardoso VR, Bravo-Merodio L, Pendleton SC, Aziz F, et al. Evaluating the detection ability of a range of epistasis detection methods on simulated data for pure and impure epistatic models. *PLoS ONE*. 2022;17:e0263390.
9. Mäki-Tanila A, Hill WG. Influence of gene interaction on complex trait variation with multilocus models. *Genetics*. 2014;198:355–67.
10. Cordell HJ. Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. *Hum Mol Genet*. 2002;11:2463–8.
11. Epistasis and evolution. *Evolutionary biology*. Oxford University Press; 2021. Available from: <https://oxfordbibliographies.com/view/document/obo-9780199941728/obo-9780199941728-0137.xml>.

12. Poelwijk FJ, Krishna V, Ranganathan R. The context-dependence of mutations: a linkage of formalisms. *PLoS Comput Biol*. 2016;12:e1004771.
13. Wan X, Yang C, Yang Q, Xue H, Fan X, Tang NLS, et al. BOOST: a fast approach to detecting gene-gene interactions in genome-wide case-control studies. *Am J Hum Genet*. 2010;87:325–40.
14. Bayat A, Hosking B, Jain Y, Hosking C, Kodikara M, Reti D, et al. Fast and accurate exhaustive higher-order epistasis search with BitEpi. *Sci Rep*. 2021;11:1–12.
15. Ponte-Fernández C, González-Domínguez J, Martín MJ. Fiuncho: a program for any-order epistasis detection in CPU clusters. *J Supercomput*. 2022;78:15338–57.
16. Balvert M. Iterative rule extension for logic analysis of data: an MILP-based heuristic to derive interpretable binary classifiers from large data sets. *INFORMS J Comput*. 2024. Available from: <https://doi.org/10.1287/ijoc.2021.0284>.
17. Pattin KA, White BC, Barney N, Gui J, Nelson HH, Kelsey KT, et al. A computationally efficient hypothesis testing method for epistasis analysis using multifactor dimensionality reduction. *Genet Epidemiol*. 2009;33:87–94.
18. Aghazadeh A, Nisonoff H, Ocal O, Brookes DH, Huang Y, Koyluoglu OO, et al. Epistatic Net allows the sparse spectral regularization of deep neural networks for inferring fitness functions. *Nat Commun*. 2021;12:5225.
19. Motsinger-Reif AA, Fanelli TJ, Davis AC, Ritchie MD. Power of grammatical evolution neural networks to detect gene-gene interactions in the presence of error. *BMC Res Notes*. 2008;1:65.
20. Li X, Liu L, Zhou J, Wang C. Heterogeneity analysis and diagnosis of complex diseases based on deep learning method. *Sci Rep*. 2018;8:6155.
21. Slim L, Chatelain C, Azencott C-A, Vert J-P. Novel methods for epistasis detection in genome-wide association studies. *PLoS ONE*. 2020;15:e0242927.
22. Chang YC, Wu JT, Hong MY, Tung YA, Hsieh PH, Yee SW, et al. GenEpi: gene-based epistasis discovery using machine learning. *BMC Bioinformatics*. 2020;21:68.
23. Knijnenburg TA, Klau GW, Iorio F, Garnett MJ, McDermott U, Shmulevich I, et al. Logic models to predict continuous outputs based on binary inputs with an application to personalized cancer therapy. *Sci Rep*. 2016;6:1–14.
24. Sun Y, Gu Y, Ren Q, Li Y, Shang J, Liu JX, et al. MDSN: a module detection method for identifying high-order epistatic interactions. *Genes*. 2022;13. Available from: <https://doi.org/10.3390/genes13122403>.
25. Weinreich DM, Lan Y, Jaffe J, Heckendorn RB. The influence of higher-order epistasis on biological fitness landscape topography. *J Stat Phys*. 2018;172:208–25.
26. Weinreich DM, Lan Y, Wylie CS, Heckendorn RB. Should evolutionary geneticists worry about higher-order epistasis? *Curr Opin Genet Dev*. 2013;23:700–7.
27. Beam AL, Motsinger-Reif A, Doyle J. Bayesian neural networks for detecting epistasis in genetic association studies. *BMC Bioinformatics*. 2014;15:368.
28. Cui T, El Mekkaoui K, Reinvall J, Havulinna AS, Marttinen P, Kaski S. Gene-gene interaction detection with deep learning. *Commun Biol*. 2022;5:1238.
29. Cybenko G. Approximation by superpositions of a sigmoidal function. *Math Control Signals Systems*. 1992;5:455–455.
30. Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. *Neural Netw*. 1989;2:359–66.
31. Lunetta KL, Hayward LB, Segal J, Van Eerdewegh P. Screening large-scale association study data: exploiting interactions using random forests. *BMC Genet*. 2004;5:32.
32. Jiang R, Tang W, Wu X, Fu W. A random forest approach to the detection of epistatic interactions in case-control studies. *BMC Bioinformatics*. 2009;10(Suppl 1):S65.
33. Yoshida M, Koike A. SNPInterForest: a new method for detecting epistatic interactions. *BMC Bioinformatics*. 2011;12:469.
34. Botta V, Louppe G, Geurts P, Wehenkel L. Exploiting SNP correlations within random forest for genome-wide association studies. *PLoS ONE*. 2014;9:e93379.
35. Holliday JA, Wang T, Aitken S. Predicting adaptive phenotypes from multilocus genotypes in Sitka spruce (*Picea sitchensis*) using random forest. *G3*. 2012;2:1085–93.
36. Li J, Horstman B, Chen Y. Detecting epistatic effects in association studies at a genomic level based on an ensemble approach. *Bioinformatics*. 2011;27:i222–9.
37. Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. *Mach Learn*. 2002;46:389–422.
38. Chen S-H, Sun J, Dimitrov L, Turner AR, Adams TS, Meyers DA, et al. A support vector machine approach for detecting gene-gene interaction. *Genet Epidemiol*. 2008;32:152–67.
39. Shen Y, Liu Z, Ott J. Support vector machines with L1 penalty for detecting gene-gene interactions. *Int J Data Min Bioinform*. 2012;6:463–70.
40. Saha S, Perrin L, Röder L, Brun C, Spinelli L. Epi-MEIF: detecting higher order epistatic interactions for complex traits using mixed effect conditional inference forests. *Nucleic Acids Res*. 2022;50:e114.
41. Cheng J, Novati G, Pan J, Bycroft C, Zemgulytė A, Applebaum T, et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science*. 2023;381:eadg7492.
42. Frazer R, Notin P, Dias M, Gomez A, Min JK, Brock K, et al. Disease variant prediction with deep generative models of evolutionary data. *Nature*. 2021;599:91–5.
43. Isacchini G, Walczak AM, Mora T, Nourmohammad A. Deep generative selection models of T and B cell receptor repertoires with soNNia. *Proc Natl Acad Sci U S A*. 2021;118. Available from: <https://doi.org/10.1073/pnas.2023141118>.
44. Lopez R, Regier J, Cole MB, Jordan MI, Yosef N. Deep generative modeling for single-cell transcriptomics. *Nat Methods*. 2018;15:1053–8.
45. Samanta B, De A, Jana G, Gomez V, Chattaraj PK, Ganguly N, et al. NEVAE: a deep generative model for molecular graphs. *J Mach Learn Res*. 2020;21:4556–88.
46. Lu Y, Lu J. A universal approximation theorem of deep neural networks for expressing probability distributions. *arXiv [cs.LG]*. 2020. Available from: <http://arxiv.org/abs/2004.08867>.

47. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. *Adv Neural Inf Process Syst.* 2014;27. Available from: [https://proceedings.neurips.cc/paper\\_files/paper/2014/hash/5ca3e9b122f61f8f06494c97b1afcc3-Abstract.html](https://proceedings.neurips.cc/paper_files/paper/2014/hash/5ca3e9b122f61f8f06494c97b1afcc3-Abstract.html).
48. Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. *Adv Neural Inf Process Syst.* 2020;33:6840–51.
49. Kingma DP, Welling M. Auto-encoding variational Bayes. *arXiv [stat.ML]*. 2013. Available from: <http://arxiv.org/abs/1312.6114v11>.
50. Liu K, Cao G, Zhou F, Liu B, Duan J, Qiu G. Towards disentangling latent space for unsupervised semantic face editing. *IEEE Trans Image Process.* 2022;31:1475–89.
51. Rives A, Meier J, Sercu T, Goyal S, Lin Z, Liu J, et al. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proc Natl Acad Sci U S A.* 2021;118. <https://doi.org/10.1073/pnas.2016239118>.
52. Schroff F, Kalenichenko D, Philbin J. FaceNet: a unified embedding for face recognition and clustering. 2015 IEEE conference on computer vision and pattern recognition (CVPR). IEEE; 2015. <https://doi.org/10.1109/cvpr.2015.7298682>.
53. Nielsen R, Akey JM, Jakobsson M, Pritchard JK, Tishkoff S, Willerslev E. Tracing the peopling of the world through genomics. *Nature.* 2017;541:302–10.
54. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature.* 2015;526:68–74.
55. Coop G. Genetic similarity versus genetic ancestry groups as sample descriptors in human genetics. *arXiv [q-bio.PE]*. 2022. Available from: <http://arxiv.org/abs/2207.11595>
56. Bhatia G, Furlotte NA, Loh PR, Liu X, Finucane HK, Gusev A, et al. Correcting subtle stratification in summary association statistics. *bioRxiv.* 2016. p. 076133. Available from: <https://www.biorxiv.org/content/10.1101/076133v1>. [cited 2024 Feb 9].
57. Sul JH, Martin LS, Eskin E. Population structure in genetic studies: confounding factors and mixed models. *PLoS Genet.* 2018;14:e1007309.
58. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet.* 2012;44:821–4.
59. Hellwege JN, Keaton JM, Giri A, Gao X, Velez Edwards DR, Edwards TL. Population stratification in genetic association studies. *Curr Protoc Hum Genet.* 2017;95:1.22.1-1.22.23.
60. Greene CS, Penrod NM, Williams SM, Moore JH. Failure to replicate a genetic association may provide important clues about genetic architecture. *PLoS ONE.* 2009;4:e5639.
61. Johnson RA, Wichern DW. *Applied multivariate statistical analysis.* London: Pearson Prentice Hall; 2007.
62. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006;38:904–9.
63. Yao Y, Ochoa A. Limitations of principal components in quantitative genetic association models for human studies. *Elife.* 2023;12. Available from: <https://doi.org/10.7554/eLife.79238>.
64. Abegaz F, Van Lishout F, Mahachie John JM, Chiachoompu K, Bhardwaj A, Duroux D, et al. Performance of model-based multifactor dimensionality reduction methods for epistasis detection by controlling population structure. *BioData Min.* 2021;14:16.
65. Ning C, Wang D, Kang H, Mrode R, Zhou L, Xu S, et al. A rapid epistatic mixed-model association analysis by linear retransformations of genomic estimated values. *Bioinformatics.* 2018;34:1817–25.
66. Lippert C, Listgarten J, Liu Y, Kadie CM, Davidson RI, Heckerman D. FaST linear mixed models for genome-wide association studies. *Nat Methods.* 2011;8:833–5.
67. Lippert C, Listgarten J, Davidson RI, Baxter S, Poon H, Kadie CM, et al. An exhaustive epistatic SNP association analysis on expanded Wellcome Trust data. *Sci Rep.* 2013;3:1099.
68. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience.* 2015;4:7.
69. Mackay TFC. Epistasis and quantitative traits: using model organisms to study gene-gene interactions. *Nat Rev Genet.* 2014;15:22–33.
70. Kuzmin E, VanderSluis B, Wang W, Tan G, Deshpande R, Chen Y, et al. Systematic analysis of complex genetic interactions. *Science.* 2018;360. Available from: <https://doi.org/10.1126/science.aao1729>.
71. Costanzo M, VanderSluis B, Koch EN, Baryshnikova A, Pons C, Tan G, et al. A global genetic interaction network maps a wiring diagram of cellular function. *Science.* 2016;353:aaf1420–aaf1420.
72. Zhao X, Li J, Liu Z, Powers S. Combinatorial CRISPR/Cas9 screening reveals epistatic networks of interacting tumor suppressor genes and therapeutic targets in human breast cancer. *Cancer Res.* 2021;81:6090–105.
73. Thompson NA, Ranzani M, van der Weyden L, Iyer V, Offord V, Droop A, et al. Combinatorial CRISPR screen identifies fitness effects of gene paralogues. *Nat Commun.* 2021;12:1302.
74. Han K, Jeng EE, Hess GT, Morgens DW, Li A, Bassik MC. Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions. *Nat Biotechnol.* 2017;35:463–74.
75. Cisneros AF, Gagnon-Arsenault I, Dubé AK, Després PC, Kumar P, Lafontaine K, et al. Epistasis between promoter activity and coding mutations shapes gene evolvability. *Sci Adv.* 2023;9:eadd9109.
76. Mapping the genetic landscape of human cells. Available from: [https://www.cell.com/cell/pdf/S0092-8674\(18\)30735-9.pdf](https://www.cell.com/cell/pdf/S0092-8674(18)30735-9.pdf).
77. Tufts DM, Natarajan C, Revsbech IG, Projecto-Garcia J, Hoffmann FG, Weber RE, et al. Epistasis constrains mutational pathways of hemoglobin adaptation in high-altitude pikas. *Mol Biol Evol.* 2015;32:287–98.
78. Rojas Echenique JI, Kryazhimskiy S, Nguyen Ba AN, Desai MM. Modular epistasis and the compensatory evolution of gene deletion mutants. *PLoS Genet.* 2019;15:e1007958.
79. Schmidt M, Lipson H. Distilling free-form natural laws from experimental data. *Science.* 2009;324:81–5.
80. Vladislavleva EY. Model-based problem solving through symbolic regression via Pareto genetic programming. CentER: Tilburg University; 2008.

81. Lundberg M, Sng LMF, Szul P, Dunne R, Bayat A, Burnham SC, et al. Novel Alzheimer's disease genes and epistasis identified using machine learning GWAS platform. *Sci Rep.* 2023;13:17662.
82. Wang H, Bennett DA, De Jager PL, Zhang QY, Zhang HY. Genome-wide epistasis analysis for Alzheimer's disease and implications for genetic risk prediction. *Alzheimers Res Ther.* 2021;13:55.
83. Park S, Lehner B. Cancer type-dependent genetic interactions between cancer driver alterations indicate plasticity of epistasis across cell types. *Mol Syst Biol.* 2015;11:824.
84. Hoffmann M, Poschenrieder JM, Incudini M, Baier S, Fritz A, Maier A, et al. Network medicine-based epistasis detection in complex diseases: ready for quantum computing. *Nucleic Acids Res.* 2024;52:10144–60.
85. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, et al. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet.* 2001;29:365–71.
86. Balvert M, Cooper-Knock J, Stamp J, Byrne RP, Mourragui S, van Gils J et al. Population structure confounds regression-based epistasis detection. 2024. Github [https://github.com/jdstamp/leiden\\_paper](https://github.com/jdstamp/leiden_paper).
87. Balvert M, Cooper-Knock J, Stamp J, Byrne RP, Mourragui S, van Gils J, et al. Population structure confounds regression-based epistasis detection. 2024. Zenodo. <https://doi.org/10.5281/zenodo.13940750>.
88. Browning SR, Browning BL, Daviglus ML, Durazo-Arvizu RA, Schneiderman N, Kaplan RC, et al. Ancestry-specific recent effective population size in the Americas. *PLoS Genet.* 2018;14(5):e1007385. <https://doi.org/10.1371/journal.pgen.1007385>.
89. Gravel S, Henn BM, Gutenkunst RN, Indap AR, Marth GT, Clark AG, et al. *Proc Natl Acad Sci U S A.* 2011;108(29):11983–8. <https://doi.org/10.1073/pnas.1019276108>.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.