# Detecting epistasis in human complex traits

Wen-Hua Wei1,2,\*, Gibran Hemani3,4,5,\*, Chris Haley1,6

1Arthritis Research UK Centre for Genetics and Genomics, Institute of Inflammation and Repair, Faculty of Medical and Human Sciences, Manchester Academic Health Science Centre, University of Manchester, Oxford Road, Manchester M13 9PT, UK

2MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK

3MRC Integrative Epidemiology Unit, University of Bristol, Oakfield House, Oakfield Grove, Bristol, BS8 2BN, UK

4Queensland Brain Institute Centre, University of Queensland, Brisbane, Queensland 4072, Australia

5University of Queensland Diamantina Institute, University of Queensland, Princess Alexandra Hospital, Brisbane, Queensland 4072, Australia

6The Roslin Institute and Royal (Dick) School of Veterinary Sciences, University of Edinburgh, Easter Bush, Midlothian, EH25 9RG, UK

\*Contributed equally

Correspondence to CSH

e-mail: Chris.Haley@igmm.ed.ac.uk

## Online summary

* Tremendous activity in the development of methodology has now rendered the exhaustive search for pairwise genetic interactions computationally routine, but addressing the statistical problems of detecting epistasis remains a big challenge.
* Most reports of epistasis influencing human complex traits that exist in the literature raise concerns regarding their validity, and do not follow the same strict protocols that are in place for reporting additive effects.
* There is mounting evidence that pairwise epistatic effects influencing human complex traits that are sufficiently large for detection in standard single-sample GWAS do not exist. If epistatic effects do influence complex traits then each interaction effect will likely be small, as is observed with additive effects.
* The majority of robust additive effects are only found when GWAS is performed using huge samples and good SNP coverage, often as a result of multi-study meta analyses. Similar approaches are necessary if epistatic effects are also to be robustly detected, though methodology or attempts at implementation are yet to surface.
* Methods have emerged for estimating the total contribution of additive effects across the whole genome, similar methods for estimation of total contribution of genetic interactions would be valuable but have yet to be developed.

## Abstract

Genome wide association studies (GWAS) have become the focus of the statistical analysis of complex traits in humans, successfully shedding light on several aspects of genetic architecture and also biological etiology. Single nucleotide polymorphisms (SNPs) are usually modeled as having additive, cumulative, and independent effects on the phenotype. Though evidently a useful approach, it is often argued that this is not a realistic biological model and that epistasis, the statistical interaction between SNPs, should be included. The purpose of this review is to summarise recent directions in methodology for detecting epistasis, and discuss evidence on the role of epistasis in human complex trait variation. We also discuss the relevance of epistasis in the context of GWAS, and potential hazards in the interpretation of statistical interaction terms.

## Introduction

**Complex traits** are those that are influenced by many genetic and environmental factors. Almost all diseases that have a significant burden on human health are complex at the population scale. Arguably the most important empirical result to emerge from GWAS over the past decade is an indication of what it means for a trait to be "complex", demonstrating that the **mutational target size** for any particular complex trait across the genome is very large, and that the additive genetic variation is comprised of very many variants, almost all of very small effect.

To date, GWAS has enjoyed most success by searching for simple additive effects, where the causal variants exhibit independent, additive and cumulative effects on the trait. But there are enduring questions about whether this body of research is revealing the full extent of the **genetic architecture** in complex traits. One aspect of this is whether polymorphisms typically act independently, or if their effects are dependent on other polymorphisms elsewhere in the genome. Defining such genetic interaction (known as **epistasis**) has been the subject of several reviews in the past. To summarize briefly, one can discuss epistasis in terms of its functional effect or its statistical effect. In terms of functional epistasis, we are referring to the general observation that the effect of a particular variant depends on the genotype of another variant. In terms of statistical epistasis, we are referring to the interaction variance explained by a combination of causal variants that is not due to their independent effects on their own. In what follows we focus largely on statistical epistasis, but we should note that the presence of functional epistasis does not automatically imply the presence of substantial statistical epistasis and *vice versa.*

There are several methods for estimating the proportion of the phenotypic variance of a trait that is attributed to independent, additive genetic effects (known as narrow-sense **heritability** (*h*2)). But owing to the technical limitations of accounting for non-additive effects, unequivocal estimation of the phenotypic variation attributable to all additive and non-additive genetic effects (known as broad-sense heritability (*H*2)) is not possible for human traits (Box 1).1 GWAS are typically performed on traits that have been shown to have a non-zero *h*2, on the assumption that this additive genetic variation in the trait can be dissected into additive effects across the genome. Searching for non-additive genetic effects imposes an assumption that, beyond the additive component of genetic variation, for which there is empirical evidence, there exist more complex components whose existence cannot be verified empirically due to technical limitations.

There is no strict hypothesis-driven precedent for searching for epistasis so why is epistasis so often at the forefront of debate? The major objectives of GWAS can be reduced down to two main categories. The first is to use knowledge of the causal variants influencing a particular trait to further understand its underlying biology. The second is to use the estimated effects of causal variants to improve prediction of phenotypic outcomes. Should the genetic architecture of a trait be comprised of substantial epistasis then identifying epistatic variants could be beneficial to both of these major objectives of GWAS: identifying instances of functional epistasis could be informative in understanding biological mechanisms; whereas should substantial levels of statistical epistasis exist then genetic prediction of complex traits could potentially be improved beyond the theoretical limit imposed by the estimate of narrow-sense heritability (Box 2).

To this end, the past few years have seen remarkable activity in the development of methodology and software for the detection of epistasis. In this review we will survey these emerging tools and follow this with a cold, hard look at the latest empirical evidence for the importance of epistasis and the potential utility in searching for genetic interactions.

## Methods for detecting epistasis

Despite the severe challenges in robust detection of epistasis for human complex traits (Box 3), the past five years have seen rapid development of methods for studying epistasis in human complex traits2–7. Methods ranging from conventional regression-based methods to nature-inspired algorithms have been developed (Figure 1) that aim to detect whether the joint effect of two or more loci differs from that predicted by their individual effects. Most methods use SNP-based tests for pairwise or higher-order interactions in GWAS data via either an **exhaustive search** of all SNP combinations or testing of a reduced, preselected set. In a pairwise exhaustive search of a GWAS cohort (typically with some thousands of samples each genotyped by < 500 000 SNPs), the primary goal is to identify interacting SNPs from a search space of many billions of pairwise tests (strictly *n*(*n*-1)/2 given *n* SNPs). Such a search creates a huge computational burden and also a major statistical challenge of achieving significance thresholds derived following **Bonferroni correction** of the number of tests performed. In addition, methods have been developed to assess interactions between groups of SNPs; for example, SNPs grouped into genes or into functional modules (pathways or networks). As a broad generalization, approaches that search all pairs of SNPs for evidence of all types of epistasis can be termed “**hypothesis free**” whereas those which restrict themselves to searching subsets of SNPs and/or types of epistasis may be termed “**hypothesis driven**”. Because of the size of the parameter space to search and the number of analyses performed, hypothesis-free approaches pose a major challenge and one of the key achievements in recent years is that computational barriers have been bypassed and an exhaustive search for pairwise interactions has become a more routine exercise8–12. Here we provide an overview of the developments in methodology since an earlier review by Cordell13.

### Regression-based methods

Traditionally methods based on logistic regression or linear regression are used to assess SNP interactions in diseases or **quantitative traits** respectively13,14 (Supplementary information box S1). Given the goal of detecting interactions, one can use SNP genotype models to test interactions directly by comparing the **saturated model** including interactions (LS) against the **reduced model** without (LR) for each pair of SNPs and thus save computing time in estimating genetic parameters unnecessarily. The (LS vs. LR) test is essentially based on the variance explained by four interaction terms and thus consumes four degrees-of-freedom (df). When concerning only additive effects, SNP allelic models can be applied instead in which case the interaction test consumes one df because only the additive-additive interaction term is included, thus this test will be more powerful if only additive effects and their interaction influence trait variation15,16.

Previously an exhaustive search for pairwise interactions at the genome-wide level was considered computationally time consuming13. Various approaches have been taken to reduce the computational barrier. First, advantage has been taken of modern computing infrastructure and technologies including clusters of computers equipped with multiple CPU cores and/or graphic processing units (GPU)9,11,12,17, parallelization8,10,18 and bitwise computing where SNP genotype data are stored in bitwise data structures to achieve great memory efficiency and computing speed8,19,20. Second, approximate interaction tests have been applied that can be quickly computed and do not miss any important epistatic SNP pairs20. For example F ratio and Kirkwood Superposition Approximation approximate the (LS vs. LR) tests under the assumption of **Hardy-Weinberg Equilibrium** (HWE) for quantitative21 and disease traits20 respectively and can be quickly computed from contingency tables based on SNP genotypes. Some recent applications based on regression and other approaches that can perform fast genome-wide screening of epistasis in GWAS are shown in TABLE 1. If using the approximate interaction tests as an initial genome-wide screen for putative epistatic interactions it is prudent to take the extra step of re-examining the resultant subset of SNP pairs using the full regression models to avoid false positives or redundant signals10,21,22.

Although computationally tractable, a genome-wide search still suffers from low power in detection of epistasis13,23,24. Large sample sizes are generally required for success25. Focusing on interactions involving SNPs with genome-wide significant **marginal effects** confirmed in GWAS may be a practical compromise26,27 that allows the identification of epistatic variance associated with loci with known individual effects, even if it will miss pairs of loci that contribute only or mainly through their interaction. Nonetheless in light of limited samples available in most individual GWAS cohorts it may be possible to identify epistasis with such an approach because a much less stringent threshold can be applied owing to much reduced multiple tests10,21,28–30. Taking 500 000 SNPs for example, the Bonferroni-corrected threshold on the -log10 scale is 12.4 for an exhaustive search in contrast to 7.0 for a hypothesis-driven search focusing on one SNP with marginal effects, equivalent to an approximate doubling of sample size in the focused search.

In future, using high-density SNPs or DNA sequence data could potentially make both exhaustive and focused interaction searches more fruitful as power is a function not only of interaction effects and sample size but also of **linkage disequilibrium** (LD) between the SNP and **causal variant** at both loci which can be maximised with very dense marker data22,25.

### LD- and haplotype-based methods**.**

In disease traits, testing interactions can be equated to testing the difference of inter-locus associations between cases and controls that may be computed faster and more powerful than logistic regression mainly because such a difference can be tested using a χ2 statistic with one df rather than the four df of logistic regression13,15,31. A collection of statistics derived to measure the inter-locus association differences between cases and controls using joint genotype frequencies generally work well for unlinked loci (e.g. on separate chromosomes) in rare (e.g. prevalence < 0.01) diseases15 (Supplementary information box S1). For example, methods based on comparing LD between pairs of SNPs in cases and controls appear to be effective in detecting interactions and an early LD-based statistic32 was recently implemented in a computationally fast tool SIXPAC for genome-wide scans of epistasis19. Haplotype-based methods adapted from the LD-based statistics can detect interactions between SNP **haplotypes** with an equal (if not better) power15. Only genotypes, rather than haplotypes, are directly observed in GWASs and thus the inference of haplotypes requires the **linkage phase** of SNP genotypes to be estimated in advance.

Caution is recommended when applying these methods in GWASs because diseases are not always rare and focusing on unlinked loci would ignore possible intra-locus interactions22,33, whereas in an exhaustive search HWE does not always hold and LD does exist for many pairs of SNPs, which could potentially generate inflated false positives15. Measures of LD based on correlations (that is, co-occurrence) between pairs of SNPs do not assume HWE or require phasing34 and have been increasingly used in studying epistasis in GWASs15,17,35. For example, for each pair of SNPs, interaction can be tested by a Z-score statistic derived from the difference of their Pearson correlation computed separately in cases and controls17,34. However, simulation studies suggest that the Z-score statistic is still subject to an inflated level of false positive results when the two SNPs are highly correlated and/or both have significant marginal effects15,16. Similarly, an improved haplotype-based statistic that incorporates a weighted average of the joint effects of two SNPs is as powerful as the Z-score statistic34 and can control false positive results at an acceptable level when only one SNP has marginal effects but not when both SNPs have marginal effects and are in LD15. This may be cured using the full logistic regression model that can also correct for **covariates** ignored in the LD and haplotype based methods16,35. Therefore a two-step solution using the Z-score statistic for screening interactions genome-wide followed by logistic regression models for testing the most promising interactions17 can control false positives without losing the power advantage.

### Bayesian methods

**Bayes’ theorem** offers a great flexibility to model and stochastically search for epistasis without enumerating all SNP combinations13. Based on a similar rationale that the difference of inter-locus genotype frequency distributions between cases and controls may indicate interactions, an early Bayesian tool “Bayesian epistasis association mapping (BEAM)”36 partitions unlinked SNPs (i.e. SNPs are physically distant) into three non-overlapping groups (i.e. unassociated, associated by marginal effects and by joint effects) according to their posterior probabilities without explicitly testing interactions7,13. The BEAM framework has been improved to use new variables accounting for LD among SNPs and thus allow a full analysis of GWAS data7,37 and further extended to be applicable for quantitative traits via two-way Bayesian partitioning (i.e. partitioning SNPs and samples simultaneously). Nevertheless, additional explicit tests for interactions among SNPs partitioned in the jointly associated group at least may be useful to exclude potential false positive interactions6. Hybrid Bayesian methods appear to be able to improve detection of epistasis in GWASs38–40; e.g. combining the strengths of the Bayesian framework and the generalized linear model allows fast and stable tests of SNP or haplotype interactions while considering covariates, marginal effects and gene–environment interactions simultaneously39,41. Additionally, the Bayesian model averaging approach may increase power of detection by averaging evidence from multiple plausible models given unknown actual interaction patterns42.

### Data-filtering methods

Hypothesis-driven approaches aim to select a subset of SNPs for interaction tests based on either existing biological knowledge (e.g. databases of pathways and protein-protein interactions5,25,43), or statistical features (e.g. marginal effects25 and SNP genotype frequencies44,45), or fast algorithms19,46–49. Methods based on **variance heterogeneity** among SNP genotypes can effectively select potentially interacting SNPs for quantitative traits; however, such methods could miss SNPs that are interacting but have limited variance heterogeneity and they require subsequent work to confirm whether gene-gene or gene-environment interactions are involved50–52. Besides the apparent speed advantage, filtering-based methods can have greater power than exhaustive searches because of much reduced multiple testing53,54, and improved functional interpretation when considering only SNPs with putative biological connections to the trait. However, caution is recommended when applying filtering because of potential biases (either upwards or downwards) caused by limitations in the algorithms and existing knowledge that may be subject to **publication bias4** and may be context-dependent55,56. Furthermore, it is debatable what threshold is appropriate after filtering which might alter the null distributions of test statistics57.

### Artificial intelligence algorithms

Many attempts have been made to adopt or improve algorithms from other disciplines to address the **large P small N problem** in detecting epistasis3,5,6,13, particularly when moving beyond pairwise interactions to consider higher-order interactions where regression-based methods3 will suffer from increasing computational complexity and data sparsity48,58. Machine-learning and data-mining algorithms often employ certain classifiers for data reduction and/or feature selection to reduce both the computational and statistical burden of an exhaustive search6,59,60, but previously were not all scaled up for GWASs or for explicitly testing interactions and could miss interactions between SNPs without marginal effects when marginal effects are emphasized in a classifier13,59,60. While most existing algorithms (e.g. **Multifactor Dimensionality Reduction**, **tree-based methods** and **entropy-based methods**) are being scaled up by modern computing technologies61–65, their classifiers are also being improved to be applicable for complex quantitative traits61,66–69, to allow the use of risk scores aggregating multiple interactions, and to account for SNPs without marginal effects70,71.

Improved computing efficiency also facilitates the combination of complementary algorithms. For example, using tree-based methods for screening and Multifactor Dimensionality Reduction for interaction testing can improve the overall performance72–74, and the Reconstructability Analysis method uses entropy-based heuristics to search and evaluate structures of various interaction models constructed by graph theory65. New algorithms such as Ant Colony Optimization, which mimics how ant colonies find the shortest route to foods75, are being adopted for epistasis studies but are not currently transparent about how interactions are tested. Nevertheless, the detection of higher-order interactions appears very challenging to machine-learning and data-mining methods too considering that interactions are not explicitly tested in many cases. When interactions are explicitly tested, e.g. using Reconstructability Analysis, detection of three-way interactions requires multilayer hypothesis tests and thus is as difficult as regression methods4,76, in addition to issues of excessive computing demand, exponentially increased multiple tests and insufficient sample sizes13. Despite some progress, it seems likely that new approaches will be needed for robust genome-wide detection of higher-order interactions.

### Group- and module-based methods

Testing interactions based on SNPs grouped into genes or functional modules can dramatically reduce the multiple test burden (e.g. only ~2 x 108 pairwise tests are required for 20,000 genes) and thus increase the power of detection24,77, including testing for higher-order interactions78. In addition, proper grouping of SNPs may collectively capture causal variants that are not well tagged by individual SNPs79. However, this may be complicated by a number of factors (e.g. group definitions and correlations among SNPs and SNP pairs)80. A common practice is to use genes to group SNPs and derive gene-based variables factoring in SNP correlations for interaction tests via regression81, or for analogous LD-based approaches15 that test the difference of correlation of a pair of genes between cases and controls82,83. Alternatively, one can first compute all pairwise interactions between SNPs in each group and derive gene-based interaction *P* values by integrating the pairwise interaction *P* values77.

These gene-based studies suggest that it is important to incorporate external LD information via **imputation** to increase the chance of capturing unobserved causal variants and to use a composite value from a set of signals (instead of the single best signal, for example) to detect multiple interactions between a pair of genes. Of note, gene-based methods implicitly assume no intragenic interactions, which may not necessarily hold22,33. One can also use pre-computed pairwise SNP interactions to build statistical interaction networks and aggregate multiple gene–gene interactions and marginal effects using network analysis algorithms such as SNPrank84 to identify functionally important pathways3. However, the issues associated with the gene-based approaches (e.g. intragenic interactions and correlated SNP pairs) also need to be addressed.

### Multi-trait and multi-level integration

Collectively analyzing multiple complementary traits may help the detection of **pleiotropic epistasis** to infer shared pathways and/or regulatory mechanisms but this domain remains fundamentally unexplored due to its potential complexity85,86. Pleiotropic epistasis could be identified in human complex traits by simply looking for SNP–SNP interactions shared across related traits22 or using a gene-based method with a latent variable representing multiple traits87. On the other hand, emerging evidence of epistasis identified through expression quantitative trait locus (**eQTL**) and other ‘omics’ studies may potentially link functional interactions and help to understand the underlying molecular mechanisms88,89. A BEAM derived Bayesian partition method that uses three latent variables for gene expression (i.e. tens of thousands traits), SNPs and individuals to model them simultaneously could be a good alternative to regression-based methods7,90. However, integration of statistical and functional interactions is not straightforward91 and may require systems biology approaches92.

### Comparing alternative analytical approaches

So far we have briefly discussed only the major methodological developments in the past five years. The methods almost all consider only genotyped SNPs with precise genotypes (i.e. they are unable to handle imputed SNPs with probability-attached genotypes at the genome-wide level) and thus are unable to support meta-analysis of multiple GWAS data sets based on imputation, which is a key future challenge to increase the power of detection of epistasis. Such a meta-analysis may also ease a common issue in contingency table based methods that lack of samples in a combined genotype would lead to reduced df, particularly when testing higher-order interactions. Another challenge is to incorporate the sex chromosome, which is currently ignored in epistasis studies93.

Obviously, it is very difficult to recommend the ‘best’ applications without careful comparisons because of huge variation among them, including interaction definitions and null distributions57. Several attempts have been made to evaluate different methods59,60,94–98, suggesting community-wide efforts99 may be essential to fulfill the task. Furthermore, consensus significance thresholds for either genome-wide or focused searches and standards for replication and reporting have become fundamentally important to improve the overall quality of future epistasis studies including result sharing (BOX 4). As a general guidance, we reiterate the recommendation of the two-step solution for a genome-wide search for epistasis, i.e. using approximate but fast interaction tests for initial screening and the full regression models to test for significance of the most promising results from the initial screen.

## Overview of empirical evidence for epistasis

The literature is replete with reports of epistasis influencing a wide spectrum of human traits and diseases, but often the evidence supporting such claims is difficult to evaluate. Different methods and experimental designs have been used, ranging from filtering approaches such as hypothesis-driven candidate gene testing, to hypothesis-free exhaustive searches. Here we provide a summary of some of these findings.

### Hypothesis-free studies

The **Wellcome Trust Case–Control Consortium** (WTCCC) data have been fruitful for identifying marginal additive effects of modest size, and indeed exhaustive two-locus searches have also been applied. Wan *et al*.20 performed a genome-wide search for pairwise interactions in each of the seven traits studied, and reported thousands of significant interactions in total. The vast majority of statistical interactions were between SNPs within the major histocompatibility complex (MHC) region affecting type 1 diabetes or rheumatoid arthritis, which may be attributable to haplotype effects, where interactions in close proximity to one another are simply tagging a single haplotype which harbors a single causal variant or perhaps several non-epistatic causal variants. It was also observed that many interactions had a specific pattern of epistasis known as multiplicative effects, where the interacting loci had marginal additive effects that were larger than expected in combination with one another. Often such an interaction can be removed simply by transforming the scale on which the trait is measured, often referred to as a scale effect. However, some of the multiplicative interactions were between SNPs that have not previously been identified by GWASs, and this lends empirical support to the idea that searching for epistasis may confer increased statistical power to detect marginal effects. Such a situation can arise if two variants have additive effects that are too small to be captured by a standard GWAS, but they have a large interaction term which can be captured when searching for interactions100. Using a larger cohort of control samples, a reduced statistical model that searched for only additive x additive interactions, and more stringent controls for population stratification, Lippert *et al*.101 also performed exhaustive scans for epistasis in the seven diseases in WTCCC. Their results largely echoed those presented in Wan *et al*., reporting very many effects in the MHC for autoimmune traits. But again there was no attempt at replication to verify these statistical claims, nor to explore the possibility of haplotype effects.

Prabhu and Pe’er used their computationally efficient software, SIXPAC, to search for epistasis influencing bipolar disorder, and identified a pair of interacting SNPs that had not previously been shown to have an effect from GWASs19. They attempted to replicate these results in an independent cohort and although the interacting regions showed some evidence for replication, the actual discovery SNPs did not.

One means to maximize detection power is to choose traits for which genetic effects are expected to be large, such as is the case in gene expression102. Using the Brisbane Systems Genetics Study103 (BSGS) data of 846 individuals with gene expression levels measured in whole blood, 501 instances of epistatic effects were detected using an exhaustive pairwise search method, of which 30 could be significantly replicated in two independent samples104. However, further analysis suggests that large, unobserved causal variants can drive the appearance of epistatic signals even after attempting to avoid this problem by filtering on LD between interacting SNPs. An important conclusion from the study was that even after correcting for power discrepancy, substantially more phenotypic variance was attributable to additive effects than to non-additive effects.

The trend that emerges is that there are hints of epistasis being uncovered through exhaustive searches for epistasis between pairs of SNPs underlying complex traits, but as of yet there is rather little evidence that this approach detects epistatic interactions that are easily interpreted and statistically replicated in comparison to additive effects. It is reasonable to conclude at this stage that large epistatic effects (that have thus far eluded detection) are unlikely to exist.

### Hypothesis-driven studies

An alternative approach to performing exhaustive searches is to overcome the problem of having a very stringent significance threshold by restricting the search to a few candidate loci. Candidate loci are typically chosen because either they are suspected to have a biological role in the trait of interest, or because they have statistical evidence for additive effects from previous GWASs. Arising from these strategies are some examples of epistasis with strong statistical support (Box 4), many examples with weak statistical support, and a few reports that declare no evidence for epistasis.

Reports for putative interactions in studies driven by biological hypotheses are rife. The sheer volume of reports of epistasis is exemplified by an important study by Combarros *et al*.105, where they collated data from over 100 publications that reported epistasis of some form influencing the risk of Alzheimer’s disease (and related traits such as its age of onset). These reports comprised multiple experimental designs, sample sizes, and statistical methods. They demonstrate that, using a standardized statistical test, 27 of the putative pairs of SNPs had interaction terms at the nominal significance level of *P* < 0.05 (mostly involving the pathogenic ε4 allele of apolipoprotein E (*APOE4*)). Although, ostensibly, this is more than what is expected by chance, Combarros *et al*. pointed out that there are many limitations to how these studies were conducted; notably, most of the studies did not adjust for relevant covariates or potential confounding from population stratification, nor did they show evidence of replication.

Owing to the variability in design and reproducibility of epistasis studies, the Epistasis Project was created to provide a regularised framework for replication of claims of epistasis in a large, well-controlled study. To date, a few reports of epistasis from candidate gene studies have been discovered106–108 or shown to replicate109 in the Epistasis Project cohort, but perhaps the main conclusion from this work is that epistasis must be cautiously reported and interpreted because in the absence of solid replication the majority of identified epistatic interactions might be false positives.

Beyond the Epistasis Project other statistically robust examples of epistasis have also been shown. For example, Rhinn *et al.*110 designed a study to identify differential gene expression caused by *APOE4* and independent of *APOE4* in conferring a risk for late-onset Alzheimer’s disease. In doing so, they demonstrated two genetic interactions where SNPs regulating expression levels of *FYN* and *RNF219* each decreased the risk of Alzheimer’s disease in *APOE4* non-carriers, but not in *APOE4* carriers. This finding was also statistically replicated in independent samples. The successful strategy of restricting the search to genetic effects that control **endophenotypes** resonates with the outcome from hypothesis-free studies, supporting the notion that examples of functional epistasis exhibit relatively little non-additive variance, thus making them hard to identify in highly polygenic complex traits.

Multiple sclerosis (MS) is another complex trait in which epistasis has been demonstrated to have an impact. Because the *HLA-DR2* haplotype in the MHC region showed complete LD over a long distance in multiple ethnic populations it was hypothesised that selection was maintaining the co-segregation of two alleles due to epistasis. *In vivo* studies in humanised mice confirmed that separation of the two alleles led to an MS-like phenotype111, and subsequent analysis in human populations showed increased incidence of MS amongst individuals exhibiting recombination between the interacting loci112; both studies thus provide support for a role for this allelic pair in protecting from MS.

The pattern by which genetic variants interact can be informative about biological function. For example, suppose an interaction is detected for a disease for which risk is only conferred at the first polymorphism in the presence of the risk allele at the second polymorphism (for robust empirical examples of this pattern see 26,27,110). This may signify that there is pathway redundancy, and each variant affects independent pathways. An alternative pattern to the one described above, where the risk allele at one locus only has an effect in the absence of the risk allele at another locus, might suggest that both variants are involved in the same pathway because the loss of either variant is sufficient to confer the effect of the loss of the pathway. A potential example of this was shown in an interaction for systemic lupus erythematosus, where not only did the interaction replicate, but the proteins encoded by the genes involved (*BLK* and *BANK1*) were shown to co-localise *in vivo*.113

An alternative strategy for narrowing the search to overcome stringent significance thresholds is to only test for epistasis amongst SNPs that have known marginal effects. Though not routine, many GWASs have reported follow-up analysis of epistasis amongst their hits, but although the number of positive findings remains very low, some successes have been reported. Strange *et al*.27 looked for epistasis amongst significant marginal effects from a GWAS for psoriasis, and demonstrated that the risk alleles at the *HLA-C* and *ERAP1* loci only conferred effects if they were both present. A similar pattern of epistasis was uncovered using the same strategy by Evans *et al*.26 in a GWAS for ankylosing spondylitis, this time between *ERAP1* and a locus at *HLA-B27* that independently exhibits a large additive effect. In both cases, the same patterns replicated in independent samples, and these are perhaps the first statistically robust examples of epistasis influencing human complex traits.

One pitfall of these examples of epistasis is that they are evident on the **observed scale** of the disease trait, but because the interaction is between SNPs with large effects, it could be the case that on the **liability scale** of disease the contribution to risk is purely additive. For example, suppose that a disease manifests only once a certain threshold of deleterious effects are present. Even if each causal variant contributes an additive risk, an individual homozygous for risk alleles at two loci with large effects might reach that threshold, whereas individuals with any other combination of alleles at these two loci will not, giving the appearance of epistasis. This is indeed a philosophical quandary when dealing with **binary phenotypes**, because on the observed scale one is implicitly using an epistatic threshold model, in the sense that the contribution of a genetic effect to an individual’s disease risk depends on the sum of risk alleles elsewhere in the genome114. Theory demonstrates that this effect is more likely to be observed when disease prevalence in the population is low, as is the case for most complex diseases.

It should be noted that although epistatic signals have been uncovered when candidate loci are selected based on known marginal effects, often this strategy is not successful. For example Lucas *et al*.115 restricted the search for epistatic effects influencing risk of myocardial infarction by hypothesizing that interactions would involve either known risk factors or SNPs that had weak marginal effects. Although the sample size was reasonably large and the search space was drastically reduced, no statistically significant epistatic signals were uncovered. Similar conclusions have been drawn for other complex traits, including type 2 diabetes116, body mass index (BMI)117 and serum uric acid levels118.

## Conclusions and future perspectives

Although plenty remains to be done, a massive body of scientific discovery has been achieved through the paradigm of additive genetic effects119 in spite of its simplicity and seeming disconnection from realistic biological models. By stark contrast, searching for epistasis has contributed rather little to the understanding of complex traits, save for the important observation that large interaction terms are very unlikely to exist between pairwise SNPs. Those areas in which searching for epistasis could facilitate the genetic analysis of complex traits have yet to receive any major dividends.

We have only focused on a few examples here, but the emerging picture from the literature is that compelling statistical support is absent for the vast majority of reported epistatic interactions. Of those interactions that have strong statistical support, the relative magnitude of statistical epistasis, and consequently its contribution to the variance of complex traits, is small. As a consequence, any improvements in prediction of complex traits from the inclusion of epistasis will probably be small also (Box 2). Potentially through some level of ascertainment bias, those epistatic effects that appear to be statistically robust predominantly have large marginal effects, and contribute relatively little in terms of phenotypic variance from interaction terms alone.

On the other hand, although there appears to be relatively little statistical epistasis, the potential for functional epistasis to shape our understanding of human biology remains. Because at this stage it appears that filtering or candidate-gene approaches are often more likely to lead to the discovery of epistasis than hypothesis-free approaches, so far biological understanding has led to the identification of epistasis (e.g. 111,120,121) more than epistasis has led to an improved biological understanding. But evidence is now emerging that hypothesis-free searches can identify epistasis104, and the search for epistasis between pairs of loci is fast becoming a relatively effortless one. Sophisticated computational techniques have made the analysis fast, interpretable, and potentially routine at the individual GWAS level. Progress in detecting epistasis in complex traits using this paradigm is likely to continue through increased sample sizes, increased SNP density and rigorous standards of reporting (Box 4).

The next challenge is to greatly improve power of detection for exhaustive pair-wise epistasis via meta-analyses of multiple GWASs (the data for which is already readily available) and to consider how to move beyond epistasis between locus pairs to capture multi-locus epistatic variance. Given the low cost of performing new analyses on pre-existing data relative to the potential benefits that could arise from knowledge of epistatic interactions, we believe that the continued search for epistasis is warranted.

## Box 1: The missing heritability

The problem of the ‘missing heritability’ refers to the observation that genetic effects discovered by GWAS do not sum to the estimate of the heritability of the trait. The metric of missing heritability emerges from two variables: the tally of genetic effects uncovered and the estimate of the trait's h2. In each case epistasis may have a subtle part to play. Firstly, some theoretical studies postulate that if, for some interacting loci, the epistatic variance is large compared to the additive variance, searching for epistatic effects might be one way to uncover loci with additive variance with more power than searching for the additive effects directly.122 Such instances might arise if the trait has a large **mutational target size** and is under direct selection.100 This scenario does not posit that epistatic variance contributes to additive variance, rather, searching for epistasis may lead to the identification of variants that are either rare or have small additive effects that would otherwise go undetected in a standard GWAS. Some empirical evidence exists to support this hypothesis,20,22 however it is unlikely to be a widespread phenomenon.123 Theory shows that even in complex multi-locus epistatic patterns, marginal additive effects are necessary to produce measurable additive variance for a trait when allele frequencies are intermediate124, and under a neutral model where allele frequencies are widely dispersed most higher order epistatic effects will appear additive.125

A second issue is that direct estimation of non-additive genetic variation is almost always intractable. This problem arises because in pedigree based studies non-additive genetic components are confounded with other components such as maternal effects, and in population based studies coefficients of kinship for non-additive effects have very small variance leading to low power to make any useful estimation without extremely large sample sizes. Consequently its contribution to the resemblance between relatives is unknown.1 Thus it is possible that under certain experimental designs heritability estimates are inflated through contamination from non-additive variation.126 To what extent is this realistically a problem and how reliably can it be measured? A recent theoretical study gained much attention after demonstrating that redundancy amongst biological pathways could create an illusion of additive genetic variance in twin studies24 (provided that large common environmental effects were present127). Indeed it is a known issue in such study designs that, although it is not the most parsimonious model, a combination of non-additive variation and common environmental variation can lead to a significant additive parameter.128

To overcome this problem one can attempt to use family-based studies to estimate additive effects directly by contrasting results from different types of relatives. For example, full siblings will share 0.5 additive variance and 0.25 dominance variance, while parent-offspring will share 0.5 additive and 0 dominance variance. If the correlation between the degree of shared additive variation and the phenotypic similarity is high across all types of relatives then this would be strong evidence for heritability estimates being uncontaminated by non-additive variance. Height shows consistent estimates of heritability between twin studies and family studies,129 but h2 estimates for BMI are much higher from twin studies than from family studies,130 suggesting that height is probably mostly influenced by additive effects but there is the potential for non-additive effects to play an important role in BMI.

## Box 2: Genetic prediction

A potential direct translation of large-scale genetic studies into medical practice is to use knowledge of an individual’s genetic profile to predict phenotypic outcomes, such as risks of future disease occurrences131. There are two main limiting factors in the accuracy of predicting phenotypes through genetic profiles. First, the maximum prediction accuracy is limited by the trait heritability.132 And second, the accuracy by which the genetic profile predicts the phenotype is dependent on the accurate estimation of the underlying genetic effects.133 In this context, would the inclusion of epistatic effects into the equivalent of a polygenc score be useful for improving genetic prediction in complex traits?

With estimates of narrow-sense heritability obtainable for complex traits, one can estimate the maximum prediction accuracy under the assumption of perfect knowledge of all additive genetic effects. Conversely, our inability to estimate broad-sense heritability therefore makes it difficult to quantify the potential improvements that the inclusion of non-additive effects might accrue. For highly heritable traits it is unlikely that non-additive genetic variance will form a large component of phenotypic variance, and therefore the inclusion of epistatic effects in genetic profile scores will not improve prediction accuracy. However one might speculate that non-additive variance could have a significant influence on the phenotypic outcome of more lowly heritable traits, and this may include endophenotypes such as DNA methylation134,135 or gene expression levels.103

Another limiting factor in the inclusion of epistatic effects for genetic prediction is that, even assuming a significant non-additive genetic component, it is demonstrably hard to obtain reliable estimates of epistatic genetic effects. Nevertheless, it is not necessary to construct genetic predictors from significant effects as estimated through GWASs alone. Using a relaxed threshold for the inclusion of additive effects into a genetic predictor has been shown to improve prediction accuracy for schizophrenia,136 demonstrating that an underlying **polygenic architecture** will comprise numerous small effects that fail to reach a stringent significance threshold. Daetwyler *et al*.133 showed that in addition to heritability, the prediction accuracy is also a function of the ratio of the number of effects influencing the trait (often modeled as the number of independent markers in the genome) and the sample size.133,137 Therefore, for most complex traits that are polygenic in nature it is necessary to use extremely large sample sizes.138 In principle one could use this approach to also include epistatic effects, but because the effective number of independent pairwise genomic regions is dramatically higher139 than independent regions, a corresponding increase in sample size may be required to obtain gains in prediction accuracy equivalent to an additive model.

There has been use of epistatic effects for the prediction of complex traits in many model organism species. It was shown that including a network of epistatic effects in chicken lines140 for prediction in independent samples141 marginally improved prediction over the use of additive effects only. A number of reports have shown similar conclusions for several traits in plants142–144, whereas others demonstrate that inclusion of epistatic effects yields no improvement in prediction accuracy,145 or that additive effects alone are sufficient to explain most genetic variation.146

## Box 3: Why is epistasis theoretically difficult to detect?

Supposing that trait variance contributed by epistatic genetic effects is comprised of numerous small effects (as is the case for the additive genetic component) the statistical power to detect them is, in principle, much lower than that of detecting additive effects for a number of reasons, as described below. Note that these challenges apply particularly to humans and other outbred species and may be greatly reduced or even eliminated altogether in studies of model organisms147.

##### Linkage disequilibrium

The variance explained by a SNP detected in a GWAS is likely to be less than the variance explained by the true causal variant that is being tagged by the marker. The additive variance at the observed marker will decrease linearly with decreasing LD r2 between itself and the causal variant, where r is the correlation between the causal variant and observed SNP. Thus GWAS success is dependent upon there being high LD between causal variants and observed SNPs. However, the dominance variance at the observed SNP reduces by a factor of r4, thus the dependence on high LD between observed SNPs and unobserved causal variants is much higher when detecting dominance effects. Extending this to two loci to consider epistasis, the additive by additive (AxA) variance on average reduces by r4 across both loci, the additive by dominance (AxD) variance by r6, and the dominance by dominance (DxD) variance by r8. The consequence of these constraints is that any given SNP-genotyping microarray has substantially greater coverage of the genome when searching for additive effects than when searching for epistatic effects and indeed dominance. To overcome this problem one needs denser genotyping or high-quality imputed genotypes or sequence data to identify non-additive effects at the same power as additive effects.

##### Curse of dimensionality

The search for additive effects is performed in a single-dimensional search space; that is, our search is constrained to the number of markers in the experiment. In principle, searching for epistasis involves expanding from one dimension to two or more dimensions, thus the parameter space increases exponentially. This problem, where any signal becomes drowned out by the noise, is known as the "curse of dimensionality". There are several strategies that one can use to scan the genome for epistatic effects as discussed in the main text, but they typically involve expansion of the search space and a higher multiple testing penalty than is required for detecting non-epistatic effects. Therefore, in order to obtain the same power of detection as searching for additive effects, the sample size must increase and/or the variance explained by epistatic effects must be larger.

##### Model complexity

Modelling the additive effect of a SNP on a phenotype is achieved through a model that uses only one degree of freedom. However, for two loci there are four interaction terms (AxA, AxD, DxA, DxD), thus under the hypothesis that each of the four epistatic effects is equal in magnitude29,30,100 the model complexity for the statistical test increases. Comparing the power of detection between genetic loci of equal variance that comprise either additive effects or epistatic effects, the increased number of degrees of freedom used by the epistatic test will lead to a less-extreme *P* value despite capturing the same amount of variance as in the scenario for the additive test. The simple way to overcome this problem is to increase the sample size.

##### Replication

Replication is key to confirming the statistical veracity of associations. Relative to testing for additive effects, the increased dependence of epistasis on high LD between observed SNPs and causal variants means that there is likely to be an ascertainment bias in a discovery sample for markers in higher LD in the sample than they are in the population, simply due to sampling. Thus, taking an independent sample from the population one resamples the LD, and it is unlikely that the same ascertainment for high LD will occur. In this context we expect the replication rate of epistasis to be substantially lower than for additive effects. Performing detection and replication on very dense SNP-genotyping microarrays, imputed genotypes, or sequence data will overcome this problem.

In summary, overcoming these challenges will require very dense marker information (i.e. dense and accurate imputation or sequence information) and very large sample sizes, perhaps 10 to 100 times those required to detect equivalent additive effects. This inevitably requires the development of effective meta-analyses tools for epistasis.

## Box 4: What constitutes a significant epistatic interaction?

Evidently, detection of epistasis is extremely challenging, and few (if any) reported epistatic effects are immune to all potential pitfalls detailed below. But with sample sizes increasing rapidly, and the growing availability of high-density SNP-genotyping microarrays and DNA sequencing, we are entering an era in which detection of epistasis is becoming feasible. Here we provide some guidance for the robust identification of epistatic interactions that can be trusted by the scientific community.

##### Replication is necessary

As with additive genetic effects, following the discovery of epistatic genetic effects, the gold standard for reporting them is to show that the same SNPs replicate with the same direction of effect on the phenotype in an independent sample.

##### Sufficiently stringent significance thresholds

If independent samples are not available for replication, stringent significance thresholds are of utmost importance. Based on there being effectively around 1 million independent loci in the genome, GWASs have adopted a standard significance threshold of *P* = 5x10-8. With the emergence of rapid computational methods for performing exhaustive pairwise scans, we suggest a significance threshold of *P* = 0.05 / 0.5(1x106)2 = 1x10-13 as it has been shown that the effective number of tests in a two locus search is very close to half the square of the effective number of independent regions in the genome139.

##### Discount the possibility of scale effects

Interaction terms can appear or disappear due to non-linear transformation of the phenotype. Ideally quantitative analysis should be performed on phenotypes that are transformed to normality, and it should be demonstrated that any detected non-additive effects persist following transformation to a biologically relevant scale148. For binary phenotypes the analysis of epistasis is most convincing when shown to be present on the liability scale of risk.

##### Discount the likelihood of haplotype effects

In a typical pairwise scan, inflation of test statistics for interaction terms between neighbouring SNPs on the same chromosome can arise due to haplotype effects, and this is often inseparable from epistatic effects.22 Ideally the two tested SNPs should neither be in LD with each other nor should both be in LD with some other, potentially unobserved, variant. This latter caveat is testable by fitting fine mapped additive SNPs (from sequence data or imputed data) as covariates with the interacting SNPs, and we suggest that interacting SNPs should be in LD r2<0.1, and normalized disequilibrium D’<0.1 to reduce the possibility of haplotype effects underlying any signals.

## Figure 1

## Types of methods to detect epistasis in GWAS. Outline of different types of methods in two major groups based on SNPs and groups of SNPs respectively.

Regression

LD

Frequentist

Genome-wide

Haplotype

Partition

Hybrid

Bayesian

Knowledge

Filtering

Statistics

SNP-based

Algorithm

Machine learning

Artificial Intelligence

Data mining

Gene

Module

Group-based

**Table 1**

A partial list of publicly available applications for genome-wide analysis of epistasis in GWAS\*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Name** | **Type** | **Trait** | **Main feature** | **URL** |
| PLINK149 | Regression | both | PLINK epistasis module has been a benchmark application for new tool developments | http://pngu.mgh.harvard.edu/~purcell/plink/epi.shtml | |
| FastEpistasis11 | Regression | both | An efficient parallel extension to the PLINK epistasis module | <http://www.vital-it.ch/software/FastEpistasis/> |
| BOOST20 | Regression | dis | Bitwise implementation, approximate tests for screening and logistic regression for testing | <http://bioinformatics.ust.hk/BOOST.html> |
| epiGPU9 | Regression | qt | GPU-enhanced contingency table based approximate tests | <http://sourceforge.net/projects/epigpu/> |
| PIAM10 | Regression | dis | Multithread parallel enhanced, contingency table based approximatetests | <http://www.ihs.ac.cn/xykong/PIAM.zip> |
| eCEO18 | Regression | dis | Bitwise implementation and cloud computing enhanced Chi square tests | http://www.comp.nus.edu.sg/~wangzk/eCEO.html | |
| SNP-SNP interactions150 | Regression | dis | Logic regression based epitasis test at the gene level, based on concepts of SNP intersection and union | http://www.ualberta.ca/~yyasui/homepage.html | |
| BiForce8 | Regression | both | Bitwise and multithread implementation and approximate tests | <http://bioinfo.utu.fi/biforcetoolbox> |
| EPIBLASTER17 | LD | dis | GPU-enhanced, LD-based screening and logistic regression for testing | <http://www.mpipsykl.mpg.de/en/epiblaster/index.html> |
| SIXPAC19 | LD | dis | Bitwise implementation of Probably Approximately Complete search algorithm and LD-based test | <http://www.cs.columbia.edu/~snehitp/sixpac/> |
| iLOCi151 | LD | dis | Parallel implementation of LD-based score test | http://www4a.biotec.or.th/GI/tools/iloci | |
| SHEsisEpi152 | Haplotype | dis | GPU and multithread enhanced odds ratio tests | http://analysis2.bio-x.cn/SHEsisMain.htm | |
| GenomeMatrix153 | Haplotype |  | Haplotype-based odds ratio test for interactions | <https://sph.uth.edu/hgc/faculty/xiong/software-B.html> |
| IndOR154 | Haplotype | dis | Independence-based odds ratio tests using a biological definition of epistasis | http://www.sites.univ-rennes2.fr/laboratoire-statistique/EMILY/IndOR/ | |
| HAPAL155 | Haplotype | dis | Mapping haplotype-haplotype interactions with adaptive LASSO | http://www.stt.msu.edu/~cui/software.html | |
| BEAM series36 | Bayesian | both | Bayesian partition model considering LD | <http://sites.stat.psu.edu/~yuzhang/> |
| BhGLM156 | Bayesian | both | Bayesian hierarchical Generalized Linear Model for haplotype interactions | <http://www.ssg.uab.edu/bhglm/> |
| SNPTEST42 | Bayesian | dis | Bayesian model averaging approach to model interactions involving known risk loci | http://www.stats.ox.ac.uk/∼marchini/software/gwas/gwas.html |
| MDR series157 | Data mining | dis or both | The MDR framework combines attribute selection and classification with cross-validation for modelling interactions, with a number of derived methods including pMDR, MdrPDT, FAM-MDR, MB-MDR (R package). | <http://www.multifactordimensionalityreduction.org/>  http://ritchielab.psu.edu/ritchielab/software/mdr-downloads/  http://www.statgen.ulg.ac.be/software/FAM-MDR/FAM-MDR.zip | |
| SNPHarvester49 | Filtering | dis | Stochastic search for significant SNP groups then test interactions using logistic regression | http://bioinformatics.ust.hk/SNPHarvester.html | |
| EDCF45 | Filtering | dis | Clustering frequent genotype combinations for testing interactions | http://www.cs.ucr.edu/∼minzhux/EDCF.zip |
| Relief series158 | Filtering | dis | Using an ensemble of tuned ReliefF filters to select SNPs for interaction tests | https://code.google.com/p/ensemble-of-filters/ | |
| SNPruler159 | Machine learning | dis | Using predictive rule to infer disease-associated epistatic interactions | http://bioinformatics.ust.hk/SNPRuler.zip | |
| Random Jungle62 | Machine learning | dis | A computational and memory efficient implementation of random forest with added features to improve classification | http://imbs-luebeck.de/imbs/de/node/227 | |
| SNPInterForest160 | Machine learning | dis | An improved random forest framework able to identify interactions between SNPs without marginal effects | https://gwas.biosciencedbc.jp/SNPInterForest/index.html | |
| RAPID47 | Data mining | dis | Identify correlated SNP pairs after projecting their correlations to distance between two points in a Euclidean space | <http://bix.ucsd.edu/projects/rapid> |
| TEAM46 | Data mining | both | Using Minimum Spanning Tree incrementally updates the contingency tables for epistatic tests without scanning all individuals | <http://www.csbio.unc.edu/epistasis/> |

\*: dis: disease trait; qt: quantitative trait; both: applicable to either disease or quantitative traits

## Glossary

**Bayes’ Theorem:** A probability theory by the Reverend Thomas Bayes to calculate conditional probabilities based on prior distributions of parameters in a model and the observed experimental data.

**Binary phenotype**: Disease traits often have two major states on the **observed scale**, diseased or healthy. They may nonetheless be complex traits where transition to the disease state is influenced by continuous variation on an underlying **liability scale** for disease that is controlled by many genetic loci and environmental effects.

**Bonferroni correction:** The simplest and perhaps most conservative method to control family error rate (*α*) by correcting for the number (*n*) of independent hypothesis tests when *n* is large, i.e. the corrected threshold *Pcorrected* = *α*/*n.*

**Broad-sense heritability (H2)** is the proportion due to all genetic effects, both additive and non-additive (including both dominance and epistasis).

**Causal variant:** A genetic variant directly modifies a phenotype and/or causes a change of disease risk. Owing to the limited amount of variation interrogated by SNP genotyping microarrays, SNPs in GWASs typically merely tag the causal region rather than themselves being the causal variant.

**Complex trait**: A trait where variation between individuals is controlled by several or many genes and different environmental effects, potentially with interactions between these different effects.

**Covariate:** A variable that may confound the outcome variable of a statistical model, e.g. age is a covariate of human height.

**Endophenotype:** A heritable trait genetically correlated with a disease trait. Often it is a trait (such as the level of a metabolite or transcript) that can be measured in all individuals whether diseased or not and which potentially provides a predictor of disease status.

**Entropy-based methods:** Entropy is a key measure of uncertainty associated with a random variable in information theory. Entropy-based methods examine the information/entropy difference between different models with and without interactions to detect epistasis.

**Epistasis:** Statistical interactions between loci in their impact on a trait such that the impact of a particular single locus genotype depends on the genotype at other loci.

**eQTL:** An expression quantitative trait locus controls variation in expression of a particular gene. An eQTL may lie adjacent to the gene being controlled (*cis* acting control) or some distance away (*trans* acting).

**Exhaustive search:** A search of all possible pairwise combinations of loci for evidence of epistatic interactions.

**Genetic architecture**: The complete description of the genetic factors influencing trait variation, such as number of genetic loci, their effects, allele frequencies actions and interactions.

**Haplotype:** A combination of alleles (DNA sequences) inherited from a single parent. A haplotype can be within one locus or across multiple loci, with or without physical coupling on the DNA strand.

**Hardy–Weinberg Equilibrium**: A principle stating that allele and genotype frequencies of variants in a population will remain constant from one generation to the next in the absence of evolutionary disturbing factors such as mutation and genetic drift.

**Heritability**: The proportion of the trait variation for a particular trait in a particular population and environment that is under genetic control.

**Hypothesis driven analysis** limits the combinations of loci tested for epistasis according to some prior hypothesis (for example that only loci with a marginal effect should be tested or only those loci in a particular pathway).

**Hypothesis free analysis:** No assumption is made about the loci involved in epistasis or their effects and so all possible pairs of SNPs are tested (an **exhaustive search**).

**Imputation:** Statistical inference of unobserved SNP genotypes based on a reference panel of known haplotypes in a population (e.g. human 1000 Genomes Project). Imputation can greatly narrow down the distance between SNPs and causal variants and thus increase the power of detection of associations.

**Large P small N problem:** A statistical challenge to estimate a large number of parameters based on a small number of samples.

**Liability scale:** Unobserved underlying risk of a binary phenotype or disease, measured on a continuous scale and is likely influenced by many genetic and environmental factors

**Linkage disequilibrium (LD)**: The nonrandom association of alleles of two or more loci in a population owing to limited recombination. LD is often used to measure the relationship of genetic markers of the loci – a high LD means the markers are closely related (i.e. co-occurring) so one marker can predict the other(s). Markers in high LD with an unobserved causal variant can increase the power of detection of associations. Several methods can be used to calculate LD, of which correlation is the simplest and robust approach.

**Linkage phase:** The information of combinations DNA alleles in a diploid individual inherited from the mother or father (also known as gametic phase).

**Marginal effect:** The average effect of a locus across all other loci and environmental effects (also known as main effect).

**Multifactor Dimensionality Reduction:** A data mining algorithm that can reduce a high-dimensional multi-locus model of multifactorial classes (i.e. SNP genotype combinations) into a one-dimensional model of one variable of either high risk (potential interacting) or low risk classes based on the ratio of cases and controls in each class. The algorithm uses cross-validation iteratively to define the best classification.

**Mutational target size:** Fraction of the genome in which new mutations potentially cause variation for a trait. For most complex traits this is large suggesting many loci can potentially influence trait variation.

**Narrow-sense heritability (h2)** The proportion of variation due to the additive effects of genes.

**Observed scale:** Measurement of a binary phenotype in terms of whether the participant exhibits or does not exhibit the phenotype

**Pleiotropic epistasis:** Statistical interaction signals shared in multiple traits.

**Polygenic architecture:** A trait genetic architecture under which many genes of small effect contribute to trait variation.

**Publication bias:** A bias that arises due to only certain types of results (e.g. those that successfully reject the null hypothesis) are much more likely to be published than others, leading to a disproportionate representation in the literature.

**Quantitative traits:** Phenotypes (e.g. height) vary continuously, in contrast to qualitative traits in which phenotypes are discrete (e.g. disease or healthy).

**Saturated and reduced models:** There are nine joint genotypes for a pair of SNPs each with three genotypes (e.g. AA, Aa and aa). These can be modelled in full using nine parameters: one as the baseline (e.g. aa/aa), two for each SNP (e.g. AA/aa and Aa/aa) and four for interactions (e.g. AA/Aa, AA/AA, Aa/Aa, Aa/AA). The saturated model fits all the nine parameters, whereas the reduced model fits the first five parameters and excludes the four interaction parameters.

**Tree-based methods:** Model-free or non-parametric machine learning approaches for conducting regression and classification analyses by recursive partitioning variables into tree structures. Popular applications in epistasis studies include random forest, random jungle, classification and regression trees.

**Variance heterogeneity:** Variance of a quantitative trait may differ between the three possible genotypes of a biallelic SNP in the presence of genetic interactions, which can therefore be used to screen for potential interacting SNPs.

**Wellcome Trust Case–Control Consortium**: One of the first large collaborative genome-wide association studies including eight disease traits. This study has become a role model for subsequent studies and the data set has been subjected to additional analyses including for epistasis.

## 

## References

1. Visscher, P. M., Hill, W. G. & Wray, N. R. Heritability in the genomics era--concepts and misconceptions. *Nat. Rev. Genet.* **9,** 255–66 (2008).

2. Huang, Y., Wuchty, S. & Przytycka, T. M. eQTL epistasis – challenges and computational approaches. *Front. Genet.* **4,** (2013).

3. McKinney, B. A. & Pajewski, N. M. Six degrees of epistasis: Statistical network models for GWAS. *Front. Genet.* **2,** 109 (2011).

4. Pang, X. *et al.* A statistical procedure to map high-order epistasis for complex traits. *Br. Bioinform* (2012). doi:bbs027 [pii] 10.1093/bib/bbs027 [doi]

5. Ritchie, M. D. Using Biological Knowledge to Uncover the Mystery in the Search for Epistasis in Genome-Wide Association Studies. *Ann. Hum. Genet.* **75,** 172–182 (2011).

6. Steen, K. V. Travelling the world of gene-gene interactions. *Br. Bioinform* **13,** 1–19 (2012).

7. Zhang, Y., Jiang, B., Zhu, J. & Liu, J. S. Bayesian Models for Detecting Epistatic Interactions from Genetic Data. *Ann. Hum. Genet.* **75,** 183–193 (2011).

8. Gyenesei, A. *et al.* BiForce Toolbox: powerful high-throughput computational analysis of gene-gene interactions in genome-wide association studies. *Nucleic Acids Res* **40,** W628–32 (2012).

9. Hemani, G., Theocharidis, A., Wei, W. & Haley, C. EpiGPU: exhaustive pairwise epistasis scans parallelized on consumer level graphics cards. *Bioinformatics* **27,** 1462–5 (2011).

10. Liu, Y. *et al.* Genome-wide interaction-based association analysis identified multiple new susceptibility Loci for common diseases. *PLoS Genet* **7,** e1001338 (2011).

11. Schüpbach, T., Xenarios, I., Bergmann, S. & Kapur, K. FastEpistasis : a high performance computing solution for quantitative trait epistasis. *Bioinformatics* **26,** 1468–1469 (2010).

12. Yung, L. S., Yang, C., Wan, X. & Yu, W. GBOOST: a GPU-based tool for detecting gene-gene interactions in genome-wide case control studies. *Bioinformatics* **27,** 1309–1310 (2011).

13. Cordell, H. J. Detecting gene-gene interactions that underlie human diseases. *Nat. Rev. Genet.* **10,** 392–404 (2009). **An excellent review of methods to study epistasis in genome-wide asssociation studies of human diseases.**

14. Cordell, H. J. Epistasis: what it means, what it doesn’t mean, and statistical methods to detect it in humans. *Hum. Mol. Genet.* **11,** 2463–2468 (2002).

15. Ueki, M. & Cordell, H. J. Improved statistics for genome-wide interaction analysis. *PLoS Genet* **8,** e1002625 (2012). **A comprehensive assessment of LD/haplotype based methods for genome-wide detection of epistasis.**

16. Hu, J. K., Wang, X. & Wang, P. Testing gene-gene interactions in genome wide association studies. *Genet. Epidemiol.* **38,** 123–34 (2014).

17. Kam-Thong, T. *et al.* EPIBLASTER-fast exhaustive two-locus epistasis detection strategy using graphical processing units. *Eur. J. Hum. Genet.* **19,** 465–471 (2010).

18. Wang, Z., Wang, Y., Tan, K. L., Wong, L. & Agrawal, D. eCEO: an efficient Cloud Epistasis cOmputing model in genome-wide association study. *Bioinformatics* **27,** 1045–1051 (2011).

19. Prabhu, S. & Pe’er, I. Ultrafast genome-wide scan for SNP-SNP interactions in common complex disease. *Genome Res* **22,** 2230–2240 (2012).

20. Wan, X. *et al.* BOOST: A Fast Approach to Detecting Gene-Gene Interactions in Genome-wide Case-Control Studies. *Am. J. Hum. Genet.* **87,** 325–340 (2010).

21. Gyenesei, A., Moody, J., Semple, C. a M., Haley, C. S. & Wei, W.-H. High-throughput analysis of epistasis in genome-wide association studies with BiForce. *Bioinformatics* **28,** 1957–64 (2012).

22. Wei, W., Gyenesei, A., Semple, C. A. M. & Haley, C. S. Properties of Local Interactions and Their Potential Value in Complementing Genome-Wide Association Studies. *PLoS One* **8,** e71203 (2013).

23. Gauderman, W. J. Sample size requirements for association studies of gene-gene interaction. *Am J Epidemiol* **155,** 478–484 (2002). **An important work investigating power and sample sizes required for studying epistasis in genome-wide association studies.**

24. Zuk, O., Hechter, E., Sunyaev, S. R. & Lander, E. S. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A* **109,** 1193–1198 (2012). **An interesting theoretical exploration of how disease traits can be the sum of many lower level pathways, and how polygenic modes of inheritance may invoke high level epistasis**

25. Ma, L. *et al.* Knowledge-driven analysis identifies a gene-gene interaction affecting high-density lipoprotein cholesterol levels in multi-ethnic populations. *PLoS Genet* **8,** e1002714 (2012).

26. Evans, D. M. *et al.* Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. *Nat. Genet.* **43,** 761–767 (2011).

27. Strange, A. *et al.* A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1. *Nat. Genet.* **42,** 985–90 (2010).

28. Carlborg, O. & Haley, C. S. Epistasis: too often neglected in complex trait studies? *Nat Rev Genet* **5,** 618–625 (2004).

29. Evans, D. M., Marchini, J., Morris, A. P. & Cardon, L. R. Two-Stage Two-Locus Models in Genome-Wide Association. *PLoS Genet.* **2,** e157 (2006).

30. Marchini, J., Donnelly, P. & Cardon, L. R. Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nat. Genet.* **37,** 413–417 (2005). **An important simulation study investigating key issues in studying epistasis in genome-wide association studies.**

31. Hoh, J. & Ott, J. Mathematical multi-locus approaches to localizing complex human trait genes. *Nat Rev Genet* **4,** 701–709 (2003).

32. Zhao, J., Jin, L. & Xiong, M. Test for interaction between two unlinked loci. *Am J Hum Genet* **79,** 831–845 (2006).

33. Haig, D. Does heritability hide in epistasis between linked SNPs? *Eur J Hum Genet* **19,** 123 (2011). **An early suggestion of examining interactions between neighbouring SNPs**

34. Wellek, S. & Ziegler, A. A genotype-based approach to assessing the association between single nucleotide polymorphisms. *Hum Hered* **67,** 128–139 (2009).

35. Yuan, Z. *et al.* From Interaction to Co-Association -A Fisher r-To-z Transformation-Based Simple Statistic for Real World Genome-Wide Association Study. *PLoS One* **8,** e70774 (2013).

36. Zhang, Y. & Liu, J. S. Bayesian inference of epistatic interactions in case-control studies. *Nat. Genet.* **39,** 1167–1173 (2007).

37. Tang, W., Wu, X., Jiang, R. & Li, Y. Epistatic module detection for case-control studies: a Bayesian model with a Gibbs sampling strategy. *PLoS Genet* **5,** e1000464 (2009).

38. Chen, G. K. & Thomas, D. C. Using biological knowledge to discover higher order interactions in genetic association studies. *Genet Epidemiol* **34,** 863–878 (2010).

39. Yi, N., Kaklamani, V. G. & Pasche, B. Bayesian analysis of genetic interactions in case-control studies, with application to adiponectin genes and colorectal cancer risk. *Ann Hum Genet* **75,** 90–104 (2011).

40. Zhang, Y. A novel bayesian graphical model for genome-wide multi-SNP association mapping. *Genet. Epidemiol.* **36,** 36–47 (2012).

41. Li, J., Zhang, K. & Yi, N. A Bayesian hierarchical model for detecting haplotype-haplotype and haplotype-environment interactions in genetic association studies. *Hum Hered* **71,** 148–160 (2011).

42. Ferreira, T. & Marchini, J. Modeling interactions with known risk loci-a Bayesian model averaging approach. *Ann Hum Genet* **75,** 1–9 (2011).

43. Turner, S. D. *et al.* Knowledge-driven multi-locus analysis reveals gene-gene interactions influencing HDL cholesterol level in two independent EMR-linked biobanks. *PLoS One* **6,** e19586 (2011).

44. Ackermann, M. & Beyer, A. Systematic detection of epistatic interactions based on allele pair frequencies. *PLoS Genet* **8,** e1002463 (2012).

45. Xie, M., Li, J. & Jiang, T. Detecting genome-wide epistases based on the clustering of relatively frequent items. *Bioinformatics* **28,** 5–12 (2012).

46. Zhang, X., Huang, S., Zou, F. & Wang, W. TEAM: efficient two-locus epistasis tests in human genome-wide association study. *Bioinformatics* **26,** i217–i227 (2010).

47. Brinza, D., Schultz, M., Tesler, G. & Bafna, V. RAPID detection of gene-gene interactions in genome-wide association studies. *Bioinformatics* **26,** 2856–2862 (2010).

48. Ueki, M. & Tamiya, G. Ultrahigh-dimensional variable selection method for whole-genome gene-gene interaction analysis. *BMC Bioinformatics* **13,** 72 (2012).

49. Yang, C. *et al.* SNPHarvester: a filtering-based approach for detecting epistatic interactions in genome-wide association studies. *Bioinformatics* **25,** 504–511 (2009).

50. Shen, X., Pettersson, M., Ronnegard, L. & Carlborg, O. Inheritance beyond plain heritability: variance-controlling genes in Arabidopsis thaliana. *PLoS Genet* **8,** e1002839 (2012).

51. Ronnegard, L. & Valdar, W. Recent developments in statistical methods for detecting genetic loci affecting phenotypic variability. *BMC Genet* **13,** 63 (2012).

52. Brown, A. A. *et al.* Genetic interactions affecting human gene expression identified by variance association mapping. *Elife* 10.7554/eLife.01381 (2014). doi:http://dx.doi.org/10.7554/eLife.01381

53. Lewinger, J. P. *et al.* Efficient two-step testing of gene-gene interactions in genome-wide association studies. *Genet. Epidemiol.* **37,** 440–51 (2013).

54. Sun, X. *et al.* Analysis pipeline for the epistasis search - statistical versus biological filtering. *Front. Genet.* **5,** 106 (2014).

55. Fairfax, B. P. *et al.* Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nat Genet* **44,** 502–510 (2012).

56. Trynka, G. *et al.* Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat Genet* **45,** 124–130 (2013).

57. Yang, C. *et al.* The choice of null distributions for detecting gene-gene interactions in genome-wide association studies. *BMC Bioinformatics* **12 Suppl 1,** S26 (2011).

58. Fang, G. *et al.* High-order SNP combinations associated with complex diseases: efficient discovery, statistical power and functional interactions. *PLoS One* **7,** e33531 (2012).

59. Culverhouse, R. C. A comparison of methods sensitive to interactions with small main effects. *Genet Epidemiol* **36,** 303–311 (2012).

60. Molinaro, A. M. *et al.* Power of data mining methods to detect genetic associations and interactions. *Hum Hered* **72,** 85–97 (2011).

61. Zhu, Z. *et al.* Development of GMDR-GPU for Gene-Gene Interaction Analysis and Its Application to WTCCC GWAS Data for Type 2 Diabetes. *PLoS One* **8,** e61943 (2013).

62. Schwarz, D. F., König, I. R. & Ziegler, A. On safari to Random Jungle: a fast implementation of Random Forests for high-dimensional data. *Bioinformatics* **26,** 1752–8 (2010).

63. Knights, J., Yang, J., Chanda, P., Zhang, A. & Ramanathan, M. SYMPHONY, an information-theoretic method for gene-gene and gene-environment interaction analysis of disease syndromes. *Heredity (Edinb).* **110,** 548–559 (2013).

64. Shervais, S., Kramer, P. L., Westaway, S. K., Cox, N. J. & Zwick, M. Reconstructability analysis as a tool for identifying gene-gene interactions in studies of human diseases. *Stat Appl Genet Mol Biol* **9,** Article18 (2010).

65. Zwick, M. Reconstructability Analysis of Epistasis. *Ann. Hum. Genet.* **75,** 157–171 (2011).

66. Lishout, F. V *et al.* An efficient algorithm to perform multiple testing in epistasis screening. *BMC Bioinformatics* **14,** 138 (2013).

67. Mahachie John, J. M., Van Lishout, F. & Van Steen, K. Model-Based Multifactor Dimensionality Reduction to detect epistasis for quantitative traits in the presence of error-free and noisy data. *Eur J Hum Genet* **19,** 696–703 (2011).

68. Gui, J. *et al.* A novel survival multifactor dimensionality reduction method for detecting gene-gene interactions with application to bladder cancer prognosis. *Hum Genet* **129,** 101–110 (2011).

69. Lee, S., Kwon, M. S., Oh, J. M. & Park, T. Gene-gene interaction analysis for the survival phenotype based on the Cox model. *Bioinformatics* **28,** i582–i588 (2012).

70. Yoshida, M. & Koike, A. SNPInterForest: a new method for detecting epistatic interactions. *BMC Bioinformatics* **12,** 469 (2011).

71. Li, J., Horstman, B. & Chen, Y. Detecting epistatic effects in association studies at a genomic level based on an ensemble approach. *Bioinformatics* **27,** i222–9 (2011).

72. Motsinger-Reif, A. A., Fanelli, T. J., Davis, A. C. & Ritchie, M. D. Power of grammatical evolution neural networks to detect gene-gene interactions in the presence of error. *BMC Res Notes* **1,** 65 (2008).

73. De Lobel, L. *et al.* A screening methodology based on Random Forests to improve the detection of gene-gene interactions. *Eur J Hum Genet* **18,** 1127–1132 (2010).

74. Lin, H. Y. *et al.* TRM: a powerful two-stage machine learning approach for identifying SNP-SNP interactions. *Ann Hum Genet* **76,** 53–62 (2012).

75. Wang, Y., Liu, X., Robbins, K. & Rekaya, R. AntEpiSeeker: detecting epistatic interactions for case-control studies using a two-stage ant colony optimization algorithm. *BMC Res Notes* **3,** 117 (2010).

76. Hu, T. *et al.* An information-gain approach to detecting three-way epistatic interactions in genetic association studies. *J Am Med Inf. Assoc* (2013). doi:amiajnl-2012-001525 [pii] 10.1136/amiajnl-2012-001525 [doi]

77. Ma, L., Clark, A. G. & Keinan, A. Gene-based testing of interactions in association studies of quantitative traits. *PLoS Genet* **9,** e1003321 (2013).

78. Oh, S. *et al.* A novel method to identify high order gene-gene interactions in genome-wide association studies: gene-based MDR. *BMC Bioinformatics* **13 Suppl 9,** S5 (2012).

79. Wu, M. C. *et al.* Powerful SNP-set analysis for case-control genome-wide association studies. *Am J Hum Genet* **86,** 929–942 (2010).

80. Wu, C. & Cui, Y. Boosting signals in gene-based association studies via efficient SNP selection. *Br. Bioinform* (2013). doi:bbs087 [pii] 10.1093/bib/bbs087 [doi]

81. He, S. & Wu, Z. Gene-based Higher Criticism methods for large-scale exonic single-nucleotide polymorphism data. *BMC Proc* **5 Suppl 9,** S65 (2011).

82. Rajapakse, I., Perlman, M. D., Martin, P. J., Hansen, J. A. & Kooperberg, C. Multivariate detection of gene-gene interactions. *Genet Epidemiol* **36,** 622–630 (2012).

83. Zhang, X. *et al.* A PLSPM-based test statistic for detecting gene-gene co-association in genome-wide association study with case-control design. *PLoS One* **8,** e62129 (2013).

84. Davis, N. A., Crowe  Jr., J. E., Pajewski, N. M. & McKinney, B. A. Surfing a genetic association interaction network to identify modulators of antibody response to smallpox vaccine. *Genes Immun* **11,** 630–636 (2010).

85. Carter, G. W., Hays, M., Sherman, A. & Galitski, T. Use of pleiotropy to model genetic interactions in a population. *PLoS Genet* **8,** e1003010 (2012).

86. Snitkin, E. S. & Segre, D. Epistatic interaction maps relative to multiple metabolic phenotypes. *PLoS Genet* **7,** e1001294 (2011).

87. Li, F. *et al.* A powerful latent variable method for detecting and characterizing gene-based gene-gene interaction on multiple quantitative traits. *BMC Genet* **14,** 89 (2013).

88. Lehner, B. Molecular mechanisms of epistasis within and between genes. *Trends Genet.* **27,** 323–331 (2011). **An overview of possible molecular mechanisms that can cause epistasis and links between functional and statistical epistasis.**

89. Becker, J., Wendland, J. R., Haenisch, B., Nöthen, M. M. & Schumacher, J. A systematic eQTL study of cis-trans epistasis in 210 HapMap individuals. *Eur. J. Hum. Genet.* 97–101 (2011). doi:10.1038/ejhg.2011.156

90. Zhang, W., Zhu, J., Schadt, E. E. & Liu, J. S. A Bayesian partition method for detecting pleiotropic and epistatic eQTL modules. *PLoS Comput Biol* **6,** e1000642 (2010).

91. Lee, S. & Xing, E. P. Leveraging input and output structures for joint mapping of epistatic and marginal eQTLs. *Bioinformatics* **28,** i137–46 (2012).

92. Holzinger, E. R. *et al.* Initialization Parameter Sweep in ATHENA: Optimizing Neural Networks for Detecting Gene-Gene Interactions in the Presence of Small Main Effects. *Genet Evol Comput Conf* **12,** 203–210 (2010).

93. Wise, A. L., Gyi, L. & Manolio, T. a. eXclusion: toward integrating the X chromosome in genome-wide association analyses. *Am. J. Hum. Genet.* **92,** 643–7 (2013).

94. Chen, C. C. *et al.* Methods for identifying SNP interactions: a review on variations of Logic Regression, Random Forest and Bayesian logistic regression. *IEEE/ACM Trans Comput Biol Bioinform* **8,** 1580–1591 (2011).

95. Garcia-Magarinos, M., Lopez-de-Ullibarri, I., Cao, R. & Salas, A. Evaluating the ability of tree-based methods and logistic regression for the detection of SNP-SNP interaction. *Ann Hum Genet* **73,** 360–369 (2009).

96. Kapur, K., Schupbach, T., Xenarios, I., Kutalik, Z. & Bergmann, S. Comparison of strategies to detect epistasis from eQTL data. *PLoS One* **6,** e28415 (2011).

97. Shang, J. *et al.* Performance analysis of novel methods for detecting epistasis. *BMC Bioinformatics* **12,** 475 (2011).

98. Winham, S., Wang, C. & Motsinger-Reif, A. A. A comparison of multifactor dimensionality reduction and L1-penalized regression to identify gene-gene interactions in genetic association studies. *Stat Appl Genet Mol Biol* **10,** Article 4 (2011).

99. An, P. *et al.* The challenge of detecting epistasis (G x G interactions): Genetic Analysis Workshop 16. *Genet Epidemiol* **33 Suppl 1,** S58–67 (2009).

100. Hemani, G., Knott, S. & Haley, C. An Evolutionary Perspective on Epistasis and the Missing Heritability. *PLoS Genet.* **9,** e1003295 (2013).

101. Lippert, C. *et al.* An exhaustive epistatic SNP association analysis on expanded Wellcome Trust data. *Sci. Rep.* **3,** 1099 (2013).

102. Schadt, E. *et al.* Genetics of gene expression surveyed in maize, mouse and man. *Nature* **422,** 297–302 (2003).

103. Powell, J. E. *et al.* The Brisbane Systems Genetics Study: genetical genomics meets complex trait genetics. *PLoS One* **7,** e35430 (2012).

104. Hemani, G. *et al.* Detection and replication of epistasis influencing transcription in humans. *Nature* **10,** 249–53 (2014).

105. Combarros, O., Cortina-Borja, M., Smith, A. D. & Lehmann, D. J. Epistasis in sporadic Alzheimer’s disease. *Neurobiol. Aging* **In Press, ,**

106. Kolsch, H. *et al.* Interaction of insulin and PPAR-alpha genes in Alzheimer’s disease: the Epistasis Project. *J Neural Transm* **119,** 473–479 (2012).

107. Bullock, J. M. *et al.* Discovery by the Epistasis Project of an epistatic interaction between the GSTM3 gene and the HHEX/IDE/KIF11 locus in the risk of Alzheimer’s disease. *Neurobiol Aging* **34,** 1309 e1–7 (2013).

108. Combarros, O. *et al.* The dopamine beta-hydroxylase -1021C/T polymorphism is associated with the risk of Alzheimer’s disease in the Epistasis Project. *BMC Med Genet* **11,** 162 (2010).

109. Combarros, O. *et al.* Replication by the Epistasis Project of the interaction between the genes for IL-6 and IL-10 in the risk of Alzheimer’s disease. *J. Neuroinflammation* **6,** 22 (2009).

110. Rhinn, H. *et al.* Integrative genomics identifies APOE ε4 effectors in Alzheimer’s disease. *Nature* **500,** 45–50 (2013). **A good example of how knowledge of protein-protein interactions can lead to the identification of statistical interactions between genetic variants**

111. Gregersen, J. W. *et al.* Functional epistasis on a common MHC haplotype associated with multiple sclerosis. *Nature* **443,** 574–577 (2006).

112. Lincoln, M. R. *et al.* Epistasis among HLA-DRB1, HLA-DQA1, and HLA-DQB1 loci determines multiple sclerosis susceptibility. *Proc. Natl. Acad. Sci.* **106,** 7542–7547 (2009).

113. Castillejo-López, C. *et al.* Genetic and physical interaction of the B-cell systemic lupus erythematosus-associated genes BANK1 and BLK. *Ann. Rheum. Dis.* **71,** 136–42 (2012).

114. Dempster, E. R. & Lerner, I. M. Heritability of Threshold Characters. *Genetics* **35,** 212–36 (1950). **A clear and insightful paper that explains the concepts behind the liability scale and observed scale in binary phenotypes**

115. Subirana, I. *et al.* Hypothesis-Based Analysis of Gene-Gene Interactions and Risk of Myocardial Infarction. *PLoS One* **7,** (2012).

116. Bell, J. T. *et al.* Genome-Wide Association Scan Allowing for Epistasis in Type 2 Diabetes. *Ann. Hum. Genet.* **75,** 10–19 (2011).

117. Wei, W. H. *et al.* Genome-wide analysis of epistasis in body mass index using multiple human populations. *Eur J Hum Genet* **20,** 857–862 (2012).

118. Wei, W. *et al.* Characterisation of genome-wide association epistasis signals for serum uric acid in human population isolates. *PLoS One* **6,** e23836 (2011).

119. Visscher, P. M., Brown, M. a, McCarthy, M. I. & Yang, J. Five years of GWAS discovery. *Am. J. Hum. Genet.* **90,** 7–24 (2012).

120. Hill, L. D. *et al.* Epistasis between COMT and MTHFR in maternal-fetal dyads increases risk for preeclampsia. *PLoS One* **6,** e16681 (2011).

121. Génin, E. *et al.* Epistatic interaction between BANK1 and BLK in rheumatoid arthritis: results from a large trans-ethnic meta-analysis. *PLoS One* **8,** e61044 (2013).

122. Verhoeven, K. J. F., Casella, G. & McIntyre, L. M. Epistasis: obstacle or advantage for mapping complex traits? *PLoS One* **5,** e12264 (2010).

123. Hill, W. G., Goddard, M. E. & Visscher, P. M. Data and Theory Point to Mainly Additive Genetic Variance for Complex Traits. *PLoS Genet.* **4,** (2008). **Explores the apparent dichotomy between evidence for functional epistasis and lack of evidence for statistical epistasis , pointing out that with allele frequency distributions typical of natural populations non-additive gene action typically generates little epistatic variance**

124. Gjuvsland, a B., Vik, J. O., Woolliams, J. a & Omholt, S. W. Order-preserving principles underlying genotype-phenotype maps ensure high additive proportions of genetic variance. *J. Evol. Biol.* **24,** 2269–79 (2011).

125. Mäki-tanila, A. & Hill, W. G. Influence of gene interaction on complex trait variation with multi-locus models Running title : multi-locus epistatic variance Key words : quantitative genetics , epistasis , additive variance , multi-locus models , selection Corresponding author : Asko . 1–27 (2014).

126. Falconer, D. S. & Mackay, T. F. C. *Introduction to quantitative genetics*. (Longman, 1996).

127. Stringer, S., Derks, E., Kahn, R., Hill, W. & Wray, N. Assumptions and Properties of Limiting Pathway Models for Analysis of Epistasis in Complex Traits. *PLoS One* **8,** 1–9 (2013).

128. Evans, D. M., Gillespie, N. a & Martin, N. G. Biometrical genetics. *Biol. Psychol.* **61,** 33–51 (2002).

129. Silventoinen, K. *et al.* Heritability of adult body height: a comparative study of twin cohorts in eight countries. *Twin Res.* **6,** 399–408 (2003).

130. Elks, C. E. *et al.* Variability in the heritability of body mass index: a systematic review and meta-regression. *Front. Endocrinol. (Lausanne).* **3,** 29 (2012).

131. Hu, X. *et al.* Integrating autoimmune risk loci with gene-expression data identifies specific pathogenic immune cell subsets. *Am J Hum Genet* **89,** 496–506 (2011).

132. Wray, N. R., Yang, J., Goddard, M. E. & Visscher, P. M. The Genetic Interpretation of Area under the ROC Curve in Genomic Profiling. *PLoS Genet.* **6,** e1000864 (2010).

133. Daetwyler, H. D., Villanueva, B. & Woolliams, J. a. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One* **3,** e3395 (2008).

134. Quon, G., Lippert, C., Heckerman, D. & Listgarten, J. Patterns of methylation heritability in a genome-wide analysis of four brain regions. *Nucleic Acids Res.* **41,** 2095–104 (2013).

135. Gervin, K. *et al.* Extensive variation and low heritability of DNA methylation identified in a twin study. *Genome Res.* **21,** 1813–21 (2011).

136. Purcell, S. M. *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460,** 748–752 (2009).

137. Wray, N. R. *et al.* Pitfalls of predicting complex traits from SNPs. *Nat Rev Genet* **14,** 507–515 (2013). **Essential bedtime reading for those interested in prediction of complex disease from genetic signals - some of the pitfalls may be even more dangerous when using epistatic signals**

138. Zhou, X., Carbonetto, P. & Stephens, M. Polygenic modeling with bayesian sparse linear mixed models. *PLoS Genet.* **9,** e1003264 (2013).

139. Becker, T., Herold, C., Meesters, C., Mattheisen, M. & Baur, M. P. Significance levels in genome-wide interaction analysis (GWIA). *Ann. Hum. Genet.* **75,** 29–35 (2011).

140. Carlborg, O., Jacobsson, L., Ahgren, P., Siegel, P. & Andersson, L. Epistasis and the release of genetic variation during long-term selection. *Nat. Genet.* **38,** 418–420 (2006).

141. Álvarez-Castro, J. M., Le Rouzic, A., Andersson, L., Siegel, P. B. & Carlborg, Ö. Modelling of genetic interactions improves prediction of hybrid patterns--a case study in domestic fowl. *Genet. Res. (Camb).* **94,** 255–66 (2012).

142. Wang, D. *et al.* Prediction of genetic values of quantitative traits with epistatic effects in plant breeding populations. *Heredity (Edinb).* **109,** 313–9 (2012).

143. Dudley, J. W. & Johnson, G. R. Epistatic Models Improve Prediction of Performance in Corn. *Crop Sci.* **49,** 763 (2009).

144. Hu, Z. *et al.* Genomic value prediction for quantitative traits under the epistatic model. *BMC Genet.* **12,** 15 (2011).

145. González-Camacho, J. M. *et al.* Genome-enabled prediction of genetic values using radial basis function neural networks. *Theor. Appl. Genet.* (2012). doi:10.1007/s00122-012-1868-9

146. Buckler, E. S. *et al.* The genetic architecture of maize flowering time. *Science* **325,** 714–8 (2009).

147. Mackay, T. F. C. Epistasis and quantitative traits: Using model organisms to study gene-gene interactions. *Nat. Rev. Genet.* **15,** 22–33 (2014). **Detection of epistasis is often more tractable in model organisms, but differences in populations and genetic architecture (especially allele frequency and effect size) make it difficult to extrapolate conclusions on the import of epistasis to human populations**

148. Houle, D., Pélabon, C., Wagner, G. & Hansen, T. Measurement and meaning in biology. *Q. Rev. Biol.* **86,** 3–34 (2011). **An interesting discussion on the science of measuring things, informative when thinking about scale effects that may underlie epistatic signals**

149. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **81,** 559–575 (2007).

150. Dinu, I. *et al.* SNP-SNP interactions discovered by logic regression explain Crohn’s disease genetics. *PLoS One* **7,** e43035 (2012).

151. Piriyapongsa, J. *et al.* iLOCi: a SNP interaction prioritization technique for detecting epistasis in genome-wide association studies. *BMC Genomics* **13 Suppl 7,** S2 (2012).

152. Hu, X. *et al.* SHEsisEpi, a GPU-enhanced genome-wide SNP-SNP interaction scanning algorithm, efficiently reveals the risk genetic epistasis in bipolar disorder. *Cell Res.* **20,** 854–857 (2010).

153. Wu, X. *et al.* A novel statistic for genome-wide interaction analysis. *PLoS Genet.* **6,** (2010).

154. Emily, M. IndOR: a new statistical procedure to test for SNP-SNP epistasis in genome-wide association studies. *Stat Med* **31,** 2359–2373 (2012).

155. Li, M., Romero, R., Fu, W. J. & Cui, Y. Mapping haplotype-haplotype interactions with adaptive LASSO. *BMC Genet.* **11,** 79 (2010).

156. Yi, N., Liu, N., Zhi, D. & Li, J. Hierarchical generalized linear models for multiple groups of rare and common variants: jointly estimating group and individual-variant effects. *PLoS Genet* **7,** e1002382 (2011).

157. Winham, S. J. & Motsinger-Reif, A. A. An R package implementation of multifactor dimensionality reduction. *BioData Min* **4,** 24 (2011).

158. Yang, P., Ho, J. W., Yang, Y. H. & Zhou, B. B. Gene-gene interaction filtering with ensemble of filters. *BMC Bioinformatics* **12 Suppl 1,** S10 (2011).

159. Wan, X. *et al.* Predictive rule inference for epistatic interaction detection in genome-wide association studies. *Bioinformatics* **26,** 30–37 (2010).

160. Winham, S. J. *et al.* SNP interaction detection with Random Forests in high-dimensional genetic data. *BMC Bioinformatics* **13,** 164 (2012).