# The title

Table of Contents

The title 1

Abstract 1

Introduction 1

Methods for detecting epistasis 2

Overview of empirical evidence for epistasis influencing complex traits 2

Hypothesis-free studies 3

Hypothesis-driven studies 3

To which scientific question(s) is epistasis the answer? 5

The missing heritability? 5

Elucidating putative biological mechanisms? 6

Evolution of complex traits? 8

Genetic prediction? 9

Personalised genomics? 10

Conclusions 10

Box 1: Why is epistasis theoretically difficult to detect? 11

Box 2: What constitutes a significant epistatic interaction? 12

Glossary 13

References 13

## Abstract

Genome wide association (GWA) studies 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 linear, 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. In this review we discuss the relevance of epistasis in the context of GWA studies, recent advances in methodology, evidence of its contribution to complex traits in humans, and potential hazards in the interpretation of statistical interaction terms.

## Introduction

Complex traits or diseases are those that are influenced by multiple environmental and genetic factors. One can make the reasonable generalization that all those diseases that have a significant burden on human health are shown to be complex at the population scale. Indeed, even classically "Mendelian" diseases that tend to be very rare, such as cystic fibrosis, are at some level complex because numerous genetic effects are involved in modifying the breadth and severity of symptoms. Arguably the most important empirical result to emerge from GWA studies 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 effects, each of very small effect sizes.

There exist several methods for estimating the proportion of the phenotypic variance of a trait that is attributed to additive genetic effects (narrow-sense heritability (*h*2)). But estimating how much of the phenotypic variation is attributed to all genetic effects (including dominance and epistasis, broad-sense heritability (*H*2)) is not possible for human traits.1 GWA studies are typically performed on traits that have been shown to have a non-zero *h*2, the logic being that there is additive genetic variation in the trait and this can be dissected into additive effects across the genome. In this context, it is already known that the additive genetic component is typically large, and its architecture is complex. Detecting non-additive genetic effects imposes an assumption that beyond the additive component of genetic variation, for which there is empirical evidence, exists more complex components whose existence cannot be verified empirically due to technical limitations. The measurable gap between the variance explained by all known additive effects detected by GWA studies and *h*2 is termed the "missing heritability". The proportion of the genetic variation that might exist beyond additive variance might be termed the "unknown heritability".

In essence, there is no strict hypothesis-driven precedent for searching for epistasis as there is for additive effects, and yet the question of epistasis in complex traits is often at the forefront of debate. In this review we will survey emerging methodology for the detection of epistasis, summarize current examples of epistasis impacting human complex traits from the literature, discuss the contexts in which epistasis may or may not be important, and suggest guidelines on what constitutes credible statistical support from empirical studies and how this should be reported.

## Methods for detecting epistasis

The past five years have seen vibrant and rapid developments of methods for studying epistasis in human complex traits/diseases2–7. The wide range of methods spans from conventional regression-based methods to nature-inspired algorithms (FIG 1). Most methods use SNP-based tests for pairwise or high order interactions in GWAS data via either an exhaustive search of all SNP combinations or testing of a reduced, preselected set. In addition, methods have been developed to assess interactions between groups of SNPs (e.g. genes) or functional modules (e.g. pathway or network). One of the key recent achievements during this course is that an exhaustive search for pairwise interactions becomes a routine exercise because the computational barrier has been greatly reduced8–12. Here we provide an overview of the developments since Cordell’s review13 with a focus on genome-wide methods.

### Regression based methods

Regression-based methods are commonly used to assess SNP interactions in either diseases or quantitative traits13,14 (BOX 1). In GWAS where billions of pairwise SNP combinations need to be assessed, the primary goal is to identify interacting SNPs from the huge search space. One can use SNP genotype models to test interactions directly by comparing the saturated model including interactions (LS) against the reduced model without (LR) using four degree-of-freedom (df) and thus save computing time in estimating genetic parameters (BOX 1). When concerning only additive effects, the genotype model is reduced to an allelic (i.e. the minor alleles of each SNP) model where the interaction is tested in the same way but using only one df15. Previously using regression-based methods in exhaustive search for pairwise interactions at the genome-wide level was computationally prohibitive13.

Various approaches have been taken to reduce the computational barrier. First, taking advantage of modern computing infrastructure and technologies including clusters of computers equipped with multiple CPU cores and/or graphic processing units (GPU)9,11,12,16, parallelization8,10,17 and bitwise computing where SNP genotype data are stored in bitwise data structures to achieve great memory efficiency and computing speed8,18,19. Second, applying approximate interaction tests that can be quickly computed without missing any important epistatic SNP pairs19, e.g. F ratio and Kirkwood Superposition Approximation approximate the (LS vs. LR) tests under the Hardy-Weinberg Equilibrium (HWE) assumption for quantitative20 and disease traits19 respectively and can be quickly computed from contingency tables based on SNP genotypes. For convenience, we list some recent applications based on regression and other approaches that can perform fast genome-wide screening of epistasis in GWAS (TABLE 1). Considering varied strengths and weaknesses in these applications, we recommend an extra step of re-examining the screening results in the full regression models (e.g. conditional tests) to avoid false positives or redundant signals10,21,22.

Nevertheless, regression-based approaches are known to suffer from low power in detection of epistasis when applying a genome-wide threshold adjusted for billions of pair-wise tests13,23,24. Big sample sizes are generally required for success, e.g. meaningful contingency table based tests require each cell to have (e.g. > 10) samples25 leading to removal of low frequency SNPs that might tag some causal variants10. Focusing on interactions involving SNPs with important marginal effects may be a practical compromise26,27 in light of limited samples available in most individual GWAS cohorts, because a much less stringent threshold can be applied owing to much reduced multiple tests10,20,28–30. Using high dense SNPs can make such focused interaction search more fruitful as power is function of interaction effects and LD between the SNP and causal variant at both loci22,25.

### LD and haplotype based methods

In disease traits, methods based on the difference of inter-locus associations between cases and controls may be more powerful than the logistic regression mainly because such difference can be tested using 1 df χ2 statistics in contrast to 4 df in regression13,15,31. Intuitively, if a haplotype of two SNPs tagging causal variants of a disease with a higher frequency in cases than in controls, i.e. the inter-locus associations differ cases from controls, it can generate apparent epistasis illustratable in a genotype-phenotype map28 or a contingency table where each joint genotype is a combination of two of the four possible haplotypes32,33. Unfortunately haplotypes are not directly observed in GWAS and require linkage phase of SNP genotypes to be estimated in advance.

Assuming HWE and linkage phase known an LD-based statistics was first used to measure inter-locus associations and indeed had a power gain33, which was recently implemented in SIXPAC to perform fast pairwise genome scans18 but with several issues to be addressed to become more applicable for GWAS: HWE does not always hold15; phasing SNP genotype data is computational expensive; intra-locus interactions are possible22,34. Correlation-based measures of LD do not require HWE or phasing35 and have been increasingly used in studying epistasis in GWAS15,16,36. For example, for each pair of SNPs one can compute their Pearson’s correlations in cases and controls separately and derive a Z-score statistics based on the difference to test interactions16,35. Unfortunately, the Z-score statistics is still subject to inflated false positives when two SNPs in LD and/or with marginal effects15, which may be cured in the logistic regression model36. Therefore a two-step solution is implemented in a GPU enhanced tool EPIBLASTER16 using Z-score statistics for screening and logistic regression models for testing and thus controls false positives without losing the power advantage in the LD-based methods.

Haplotype-based methods face an additional issue: uncertainty in estimating haplotype frequencies especially when the HWE assumption does not hold15,37. An improved haplotype-based statistics that adopts correct variance calculation and incorporates a weight average of the joint effects of two SNPs is as powerful as the Z-score statistics35 and can control false positives when only one SNP has marginal effects, but still generates inflated false positive particularly if both SNPs have marginal effects and are in LD15. Arguably such inflation could be viewed as strengthen of the haplotype-based methods in terms of identifying ‘co-associated’ marginal loci without interactions36,38 if they are missed in conventional GWAS. The two-step solution16 may be an effective fix of the inflation issue here.

### Bayesian methods

Bayes’ theorem offers a great flexibility to model and stochastically search epistasis without enumerating all SNP combinations39. In BEAM the first Bayesian tool for GWAS40, detection of interacting SNPs is equivalent to partition independent SNPs (i.e. no LD) into predefined groups according to their posterior probabilities without explicitly testing interactions6, 12. Improved BEAM methods use new variables to account for LD among SNPs and thus allow a full analysis of GWAS data6, 41, but may benefit more from additional tests for interactions among SNPs partitioned in the target group5. Hybrid Bayesian methods appear to be able to improve detection of epistasis in GWAS42-44, e.g. combining the strengths of Bayesian framework and generalized linear model allows fast and stable tests of SNP or haplotype interactions while considering covariates, marginal effects and gene-environment interaction simultaneously43, 45. Besides, the Bayesian model averaging approach may increase power of detection by averaging evidence from multiple plausible models given unknown actual interaction types46.

### Data filtering methods

A given GWAS data can be filtered to select a subset for interaction tests based on either existing biological knowledge (e.g. databases of pathways and protein-protein interactions4, 26, 47), or statistical features (e.g. marginal effects26 and SNP genotype frequencies48, 49), or fast algorithms18, 50-53. Methods based on variance heterogeneity can effectively select potentially interacting SNPs for quantitative traits but could miss interacting SNPs lack of variance heterogeneity54, 55. Besides the apparent speed advantage, filtering based methods can be better than exhaustive search in power because of much reduced multiple tests as well as functional interpretation when considering only functional SNPs. However, cautions are 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 specific contexts56, 57. Furthermore, it is debatable what threshold is appropriate after filtering as it might alter the NULL distributions58.

### Machine learning, data mining and other algorithms

Continuous efforts have been made to adopt/improve algorithms from other disciplines to address the large P small N problem in detecting epistasis2, 4, 5, 12, particularly high-order interactions where regression-based methods3 may suffer from increasingly computational complexity and data sparsity52, 59. These algorithms often employ certain classifiers for data reduction and/or feature selection to reduce both the computational and statistical burden in exhaustive search60-62, but previously were not scaled up for GWAS or explicitly testing interactions and could miss interactions without marginal effects12, 61, 62. While most existing algorithms (e.g. Multifactor Dimensionality Reduction, tree-based, entropy-based) being scaled up by modern computing technologies63-67, their classifiers are also improved to be applicable for complex quantitative traits63, 68-71, allow using risk scores aggregating multiple interactions in classification72, and account for interactions without marginal effects73, 74.

Improved computing efficiency also allows an easy fix of the stability issue of the RelifF-based filtering methods75 by aggregating multiple runs76 and facilitates the development of combining complementary algorithms. For example, using tree-based methods for screening and Multifactor Dimensionality Reduction for interaction testing can improve the overall performance77-79; Reconstructability Analysis method uses entropy-based methods to construct and interpret interaction structures and graph theory heuristics to traverse67. New algorithms such as Ant Colony Optimization mimicking how ant colonies find the shortest route to foods80, 81 are continuously adopted into epistasis studies but need to be clear how interactions are tested81. Nevertheless, detection of high-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 methods3, 82, in addition to issues of excessive computing demand, exponentially increased multiple tests and insufficient sample sizes12. Novel ideas are needed to genome-wide detect high-order interactions.

### Group and module based methods

Testing interactions based on groups of SNPs or functional modules can dramatically reduce the multiple test burden (e.g. only ~108 pairwise tests required for 20,000 genes) and thus increase power of detection25, 83, including high-order interactions84. In addition, properly grouping SNPs may collectively capture casual variants that are not well tagged by individual SNPs85 but could be complicated by a number of factors (e.g. group definitions, correlations among SNPs and SNP pairs)38. A common practice is to use genes to group SNPs and derive gene-based variables factoring in SNP correlations for interaction tests via regression86 or analog LD-based14 (i.e. testing the difference of correlation of a pair of genes between cases and controls) approaches87, 88. 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 values83.

These gene-based studies suggest it is important to incorporate external LD (e.g. via imputation) to capture unobserved causal variants and use a set of signals instead of the single best to detect multiple interactions between a pair of genes. Nonetheless, the gene-based methods implicitly assume no intragenic interactions which may not necessarily hold22, 35 and could potentially jeopardize the power. One can also use pre-computed pairwise SNP interactions to build statistical interaction network and aggregate multiple gene-gene interactions and marginal effects using network analysis algorithms such as SNPrank89 to identify functional important pathways2. However, the issues associated with the gene-based approaches also need to be addressed here.

### Multi-trait and multi-level integration

Collectively analyzing multiple complementary traits may help detection pleiotropic epistasis for better biological interpretation but remain fundamentally unexplored due to the amounted complexity90, 91. 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 traits92. On the other hand, emerging evidence of epistasis from eQTL and other -omics studies can help understand the underlying molecular mechanisms93, 94. However, integration statistical and functional interactions is not straightforward95, 96 and may require system biology approaches97. Further efforts are needed to better integrate interaction signals at both directions.

So far we have briefly discussed only the major method developments in the past five years. Obviously, it is very difficult to recommend the ‘best’ applications without careful comparisons because of huge variation among them, including interaction definitions and NULL distributions58. Several attempts have been made to evaluate different methods previously61, 62, 98-102, suggesting community-wide efforts103 may be essential to fulfill the task. Furthermore, consensus significance thresholds and standards for replication and reporting have become fundamentally important to improve the overall quality of future epistasis studies including result sharing.

## Overview of empirical evidence for epistasis influencing complex traits

The literature is replete with reports of epistasis influencing human traits, but often the veracity of these claims is difficult to establish. A range of methods and experimental designs have been used, ranging from hypothesis-driven candidate gene testing, to hypothesis-free searches. And indeed epistasis has been reported for a wide spectrum of complex traits and diseases. Here we provide a summary of some of these findings.

### Hypothesis-free studies

The Welcome Trust Case Control Consortium (WTCCC) data has been fruitful in identifying **marginal** additive effects of modest size, and indeed exhaustive two-locus searches have been applied also. Wan et al19 used BOOST software to perform a search for pairwise interactions in each of the seven traits, and claimed to have identified thousands of significant interactions in total. The vast majority of interactions were between SNPs in the MHC region affecting type 1 diabetes or rheumatoid arthritis, which may be attributable to haplotype effects; and had a multiplicative effect, which may suggest a scale effect. However, some of the multiplicative interactions were between SNPs that have not previously been identified by GWAS, and this lends empirical support to the idea that searching for epistasis may confer increased statistical power to detect marginal effects. Using a different statistical model and more stringent controls for population stratification, Lippert et al40 also performed exhaustive scans for the seven diseases in WTCCC. Their results largely echoed those presented in Wan et al, but again there was no attempt at replication to verify these statistical claims.

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 GWAS. They also attempted to replicate these results, however although the interacting regions showed some evidence for replication, the actual discovery SNPs did not.

The trend that emerges is that there are hints of epistasis being uncovered through exhaustive searches, but as of yet there is no evidence that this approach detects epistatic interactions that are easily interpreted and statistically replicated. Resonating with this conclusion is a study by Lucas et al,41 which restricted the search for epistatic effects influencing risk of myocardial infarction to either SNPs that had weak marginal effects or to SNPs that had known marginal effects for a number of related traits. Though 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 too, including type 2 diabetes42, BMI43 and serum uric acid levels44.

### Hypothesis-driven studies

An alternative approach to performing exhaustive searches is to overcome the problem of having a very stringent threshold by restricting the search to a few candidate loci. Using this strategy, there are some examples of epistasis with strong statistical support, and many examples with weak statistical support in the literature.

The sheer volume of reports of epistasis is exemplified by an important study by Combarros et el45, where they collated data from over 100 publications that reported epistasis of some form influencing Alzheimer’s disease (and related traits, such as age of onset etc.). These reports comprised multiple experimental designs, sample sizes, and statistical methods, but they demonstrate that, using a standardized statistical test, 27 had interaction terms at the nominal significance level of *p < 0.05* (mostly involving APOE4). Though, 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, nor did they show evidence of replication.

The Epistasis Project was created for this reason, 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 discovered46–48 or shown to replicate49 in the Epistasis Project cohort, but perhaps the main conclusion from this work is that one must be cautious in reporting or interpreting epistasis because it appears that the majority are false positives.

Beyond the Epistasis Project other statistically robust examples of epistasis have been shown. For example, Rhinn et al50 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 FYN and RNF219 each decreased the risk of Alzheimer’s in APOE4 non-carriers, but not in APOE4 carriers. This finding was also statistically replicated in independent samples. The strategy of restricting the search to genetic effects that control **endophenotypes** that are involved in the trait of interest is an attractive idea, because it is expected that genetic effects influencing endophenotypes will be larger than those influencing higher-order traits, perhaps due to their less **polygenic architecture**.

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 linkage disequilibrium 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 phenotype51, and subsequent analysis in human populations showed increased incidence of MS amongst individuals exhibiting recombination between the interacting loci.52

An alternative strategy for narrowing the search to overcome large significance thresholds is to only test for epistasis amongst SNPs that have known marginal effects. Though not routine, many GWA studies report that they performed follow-up analysis of epistasis amongst their hits, but though the number of positive findings remains very low, some successes have been reported. Strange et al53 looked for epistasis amongst significant marginal effects from a GWA study 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 al26 in a GWA study for ankylosing spondylitis, this time between ERAP1 and a large additive effect at HLA-B27. 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 criticism 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. 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 risk of becoming affected by a disease depends on the sum of risk alleles elsewhere in the genome.54 This non-linearity is particularly elevated when prevalence in the population is low, as is the case for most complex diseases.

Given the success of identifying epistasis through endophenotypes, and by restricting the search to SNPs with known marginal effects, it is intuitive that combining these strategies should be fruitful. Becker et al55 searched for epistasis influencing gene expression, restricting the search to be for probes with known cis-acting eQTLs, and looking for interactions only between a cis-acting eQTL and all remaining trans-SNPs. Though it is unclear if any of these results surpassed a significance threshold that accounts for the experiment-wide testing burden, and no replication was attempted, it appears that there was enrichment for probes with nominally significant interactions. This is in stark contrast to the numerous additive effects that are typically observed in eQTL studies.56

Though we have only focused on a few examples here, the picture that is emerging is that compelling statistical support is absent for the vast majority of reported epistatic interactions. Potentially through some level of ascertainment bias, those epistatic effects that appear to be statistically robust have predominantly large marginal effects, and contribute relatively little in terms of phenotypic variance from interaction terms alone. Given the current empirical evidence, one can hypothesise that the number of instances of epistasis may indeed be large, but that the contribution of variance from gene interactions on complex traits is rather small.

## To which scientific question(s) is epistasis the answer?

### The missing heritability?

By definition, epistasis does not have a direct role to play in the problem of the missing heritability. The missing heritability pertains to the problem that there are insufficient additive effects discovered to account for the additive genetic variance estimated to exist for that trait. SNPs that interact can have marginal additive (and dominance) effects, but the nexus to declaring that they interact is that they have an epistatic effect that does not contribute to additive genetic variation. In principle, the total additive genetic variance accounted for by interacting SNPs can be captured by a standard, suitably powered GWA study searching for additive effects.

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.57 Such instances might arise if the trait has a large **mutational target size** and is under direct selection.58 This scenario does not posit that epistatic variance contributes to additive variance, rather, searching for epistasis may lead to the identification of variants with small additive effects that would otherwise go undetected in a standard GWAS. Some empirical evidence exists to support this hypothesis,19,22 however it is unlikely to be a widespread phenomenon59 and theory shows that even in complex multi-locus epistatic patterns marginal additive effects are necessary to produce measurable additive variance for a trait.60

Secondly, because the direct estimation of non-additive genetic variation is almost always intractable, 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.61 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 studies62 (provided that large common environmental effects were present63). Indeed it is a known issue in such study designs that though it is not the most parsimonious model, a combination of non-additive variation and common environmental variation can lead to a significant additive parameter.64

To overcome this problem one can attempt to use family-based studies to estimate additive effects directly, and one can use contrasts between 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 degree of shared additive variation and 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,65 but h2 estimates for BMI are much higher from twin studies than from family studies,66 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.

### Elucidating putative biological mechanisms?

In principle, knowledge of genetically interacting regions could be very useful for two reasons. Firstly, like any associations, the actual locations of the variants can shed light on the mechanisms underlying the trait by virtue of being close to relevant genes or genomic features, and as discussed in the previous section theory suggests that in some situations, searching for epistasis directly may improve power to detect variants with small marginal effects.58,67,68

Secondly, the pattern by which they interact genetically can also be informative. For example, suppose an interaction is detected for a disease where 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 26,27,50). 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 erythratosus, where not only did the interaction replicate, but the genes involved (BLK and BANK1) were shown to co-localise *in vivo*.69

An extension of these ideas is to deconstruct complex traits into endophenotypes70 and analyse the endophenotypes themselves. It has been shown theoretically that if the endpoint of a metabolic pathway depends on the rate of expression of different enzymes within that pathway, then even if the rate of expression of each individual enzyme was controlled by purely additive genetic effects, then in many situations the genetic effects on the pathway outcome would appear non-additive.71 Further theoretical studies have shown that epistasis will form a large component of variation in pathway endpoints if there are negative feedback loops controlling the outcome.72

The limiting pathway model62 employs this idea to postulate that if a disease state depends on the states of multiple independent endophenotypes, where each effectively endophenotype63 has a polygenic additive genetic architecture, then there will be non-additive variance contributing to the disease state. Though such a model is biologically intuitive, it inherently specifies that power improvements can be made by modeling endophenotype networks and searching for additive effects within these lower level traits,73 rather than advocating the use of epistatic models to map genetic effects.

If epistasis were easily detectable in a hypothesis-free framework, its potential to drive higher biological understanding would be very attractive. But because at this stage it appears that 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 identifying epistasis (e.g. 51,74,75) more than epistasis has led to improving biological understanding.

### Evolution of complex traits?

The evolutionary mechanisms that lead to phenotypic variation in the population are wide ranging, and a purely additive framework can be used to parsimoniously explain the extant standing variation in human populations.76–80 But there is a large body of theory that discusses the role that **functional epistasis** plays in long term evolutionary models, with some compelling evidence of its existence in model organisms.81 But functional epistasis is of relatively low importance compared to **statistical epistasis** in dealing with phenotypic variation within populations. In this context, to what extent does evidence support the notion that epistasis is an important factor,81–83 and what is the consequence of epistasis influencing the evolution of human complex traits?

It has been argued that epistasis plays a role in different responses to genetic effects between species through the mechanism of **Dobzhansky-Muller incompatibility**,84,85 where non-synonymous mutations in humans exist with no adverse phenotype in other species, demonstrating a form of functional epistasis. For example, in humans a non-synonymous mutation causing the 53rd amino acid of alpha-synuclein from alanine to threonine predisposes to Parkinson’s disease. However, the homologous site in healthy mice and rats carries threonine, implying that there exist substitutions elsewhere that compensate for the effect. Such compensatory molecules can exist on the same molecule or on an entirely different molecule, and one study found 608 such examples of genetic suppression in a survey of 32 human proteins.86

Does this same mechanism arise within populations to influence complex traits through statistical epistasis? A compelling line of evidence has been shown by Lappalainen et al87 in gene expression. Using sequence data, they showed that haplotypes carrying an allele that increased gene expression levels were significantly less likely to carry a putative functional coding variant in the gene due to allelic imbalance, suggesting that variants may be exposed to or masked from selection depending on their genetic backgrounds.58 In this case, it is clear that the evolution of complex traits are influenced by functional epistasis, but they will exhibit almost no statistical epistasis given the very low frequencies of the coding variants.

As shown in the example of the HLA-DR2 haplotype’s involvement in MS, another effect of selection acting on such local interactions is to prevent the decay of linkage disequilibrium across distant loci. It is difficult to ascertain the extent in which this occurs, because a) testing for LD being higher than expected by chance has very little power over short distances, b) it is very difficult to obtain orthogonal estimates of epistatic effects between SNPs in LD with each other,88 and c) the statistical effects across the population that would allow detection of such interactions are precisely the effects being minimized by selection in the first place. Nevertheless, it has been argued that such local interactions may indeed have an impact on the problem of the missing heritability. Haig (2011)89 theorised that in pedigree-based measures of heritability, haplotypes are essentially inherited as if they were multi-allelic blocks of DNA, such that the haplotype effect will contribute to heritability estimates, but if alternate haplotypes had opposing effects then on average each SNP will have no marginal additive effect.

### 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.90 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.91 And second, the accuracy in which the genetic profile predicts the phenotype is dependent on the accurate estimation of the underlying genetic effects.92 In this context, is the inclusion of epistatic effects into genetic scores important 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 incur. 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 methylation93,94 or gene expression levels.95

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 GWA studies alone. Using a relaxed threshold for the inclusion of additive effects into a genetic predictor has been shown to improve prediction accuracy for schizophrenia,96 demonstrating that an underlying polygenic architecture will comprise numerous small effects that fail reach a stringent significance threshold. Daetwyler et al92 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.92,97 Therefore, for most complex traits that are polygenic in nature it is necessary to use extremely large sample sizes.98 In principle one could use this approach to also include epistatic effects, but because the effective number of independent pairwise genomic regions is dramatically higher99 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 extensive 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 lines100 for prediction in independent samples101 marginally improved prediction over the use of additive effects only. A number of reports have shown similar conclusions for several traits in plants also102–104, whereas others demonstrate that inclusion of epistatic effects yields no improvement in prediction accuracy,105 or that additive effects alone are sufficient to explain most genetic variation.106

### Personalised genomics?

From a population perspective, the importance of a particular causal variant is proportional to its contribution to phenotypic variance. Thus far, there is very little evidence that epistasis plays a substantial role in this regard. From a personalized genomics perspective, however, the main emphasis lies on the effect a causal variant has on an individual’s phenotype regardless of population effects. In this regard, there is a growing body of evidence to support the notion that functional epistasis is indeed widespread across the genome, and impacting numerous complex traits.81,107

A recent study has shown that using exome sequencing on patients exhibiting a range of disease phenotypes can lead to the identification of rare, putative underlying causal mutations in 25% of cases.108 Another study reported that 3.4% Europeans are likely to have at least one high penetrance pathogenic or likely-pathogenic variant that for which there is potential intervention treatment available.109 The extent to which functional epistasis a) exists compared to mutations acting independently, b) is detectable, and c) is likely to play an important role in improving medical care are questions that cannot yet be answered. We are still in the early stages of translating knowledge of genetic variation into actionable medical practice110, but give the greater evidence for functional epistasis compared to statistical epistasis, perhaps personalized genomics is an area in which epistasis will play a larger role in the analysis of human complex traits.

## Conclusions

Though plenty remains to be done, a massive body of scientific discovery has been achieved through the additive genetic paradigm,111 in spite of its simplicity and seeming disconnection from realistic biological models. With a clear view on how to continue this progress (increase sample sizes, increase SNP density), it is not clear that epistasis is offering a solution to the major questions being posed about the genetics of human complex traits.

Nevertheless, the search for epistasis is fast becoming a relatively effortless one. Sophisticated computational techniques have made the analysis fast, interpretable, and potentially routine. GWAS sample sizes are constantly growing whether epistasis is considered or not. Therefore we believe that given this low cost to high potential benefit scenario the search for epistasis is indeed warranted.

## Box 1: Why is epistasis theoretically difficult to detect?

Supposing that non-additive variance 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.

##### Linkage disequilibrium

The variance explained by a SNP detected in a GWAS is unlikely to be equal to 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, thus if effect sizes are small then GWAS 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, the AxA variance on average reduces by r4 across both loci, the AxD variance by r6, and the AxD variance by r8. The consequence of these constraints is that any given SNP chip has substantially greater coverage of the genome when searching for additive effects than when searching for epistatic effects. To overcome this problem one needs denser genotyping, 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, 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 one depends on the epistatic variance being 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 magnitude29,58,112 the model complexity for the statistical test increases. In hypothetically comparing the power to 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 sample size.

##### Replication

Replication is key to confirming the statistical veracity of associations. The increased dependence of epistasis on high LD between observed SNPs and causal variants means that there is likely to be ascertainment 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 it is 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 chips, imputed genotypes, or sequence data will overcome this problem.

## Box 2: What constitutes a significant epistatic interaction?

Evidently, detection of epistasis is extremely challenging. But with sample sizes increasing rapidly, and the growing availability of high density SNP chips and sequencing, we are entering an era in which detection of epistasis may indeed be feasible. Here we suggest some considerations in how to report epistatic interactions.

##### 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 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 approximately 1 million independent regions in the genome, GWAS studies 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.

##### Discount the possibility of scale effects

Ideally quantitative analysis should be performed on phenotypes that are normalized through rank transformation, and it should be demonstrated that any detected non-additive effects persist following transformation of to a biologically relevant scale.113 For case-control traits 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 We suggest that interacting SNPs should be in LD r2<0.1, and D’<0.1 to reduce the possibility of haplotype effects underlying any signals.

## Glossary

Complex traits

Mutational target size

Additive genetic variance

Marginal effects

Endophenotype

eQTL

## 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).

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).

16. Kam-Thong, T. *et al.* EPIBLASTER-fast exhaustive two-locus epistasis detection strategy using graphical processing units. *Eur J Hum Genet* **19,** 465–471 (2011).

17. 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).

18. Prabhu, S. & Pe’er, I. Ultrafast genome-wide scan for SNP-SNP interactions in common complex disease. *Genome Res.* **22,** 2230–40 (2012).

19. 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).

20. Gyenesei, A., Moody, J., Semple, C. A., Haley, C. S. & Wei, W. H. High-throughput analysis of epistasis in genome-wide association studies with BiForce. *Bioinformatics* **28,** 1957–1964 (2012).

21. Gyenesei, A., Semple, C. A. M., Haley, C. S. & Wei, W.-H. Corrigendum of “High throughput analysis of epistasis in genome-wide association studies with BiForce.”*Bioinformatics* (2013). at <http://bioinformatics.oxfordjournals.org/content/early/2013/08/31/bioinformatics.btt444.abstract>

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).

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).

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,** 985–990 (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).

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

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

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

34. Haig, D. Does heritability hide in epistasis between linked SNPs? *Eur J Hum Genet* **19,** 123 (2011).

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

36. 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).

37. 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]

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

39. Cordell, H. J. Detecting gene-gene interactions that underlie human diseases. *Nat Rev Genet* **10,** 392–404 (2009).

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

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

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

43. 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).

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

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

46. 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).

47. 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).

48. 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).

49. 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).

50. Rhinn, H. *et al.* Integrative genomics identifies APOE ε4 effectors in Alzheimer’s disease. *Nature* **500,** 45–50 (2013).

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

52. 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).

53. 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).

54. Dempster, E. R. & Lerner, I. M. Heritability of Threshold Characters. *Genetics* **35,** 212–36 (1950).

55. Becker, J., Wendland, J. R., Haenisch, B., Nothen, M. M. & Schumacher, J. A systematic eQTL study of cis-trans epistasis in 210 HapMap individuals. *Eur J Hum Genet* **20,** 97–101 (2012).

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

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

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

59. 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).

60. 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).

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

62. Zuk, O., Hechter, E., Sunyaev, S. R. & Lander, E. S. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc. Natl. Acad. Sci.* (2012). doi:10.1073/pnas.1119675109

63. 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).

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

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

66. 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).

67. Greene, C. S., Penrod, N. M., Williams, S. M. & Moore, J. H. Failure to Replicate a Genetic Association May Provide Important Clues About Genetic Architecture. *PLoS One* **4,** e5639 (2009).

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

69. 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).

70. Gottesman, I. I., Ph, D. & Gould, T. D. Reviews and Overviews The Endophenotype Concept in Psychiatry : Etymology and Strategic Intentions. 636–645 (2003).

71. Keightley, P. D. Models of quantitative variation of flux in metabolic pathways. *Genetics* **121,** 869–76 (1989).

72. Gjuvsland, A. B., Hayes, B. J., Omholt, S. W. & Carlborg, O. Statistical epistasis is a generic feature of gene regulatory networks. *Genetics* **175,** 411–20 (2007).

73. Houle, D., Govindaraju, D. R. & Omholt, S. Phenomics: the next challenge. *Nat. Rev. Genet.* **11,** 855–866 (2010).

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

75. 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).

76. Fisher, R. The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinburgh* **52,** 399–433 (1918).

77. Hill, W. G. Rates of change in quantitative traits from fixation of new mutations. *Proc. Natl. Acad. Sci. U. S. A.* **79,** 142 (1982).

78. Eyre-Walker, A. Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. *Proc. Natl. Acad. Sci. U. S. A.* **107 Suppl ,** 1752–6 (2010).

79. Barton, N. H. & Keightley, P. D. Understanding quantitative genetic variation. *Nat. Rev. Genet.* **3,** 11–21 (2002).

80. Crow, J. F. On epistasis: why it is unimportant in polygenic directional selection. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **365,** 1241–4 (2010).

81. Phillips, P. C. Epistasis--the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet* **9,** 855–867 (2008).

82. Hansen, T. F. & Wagner, G. P. Modeling genetic architecture: a multilinear theory of gene interaction. *Theor. Popul. Biol.* **59,** 61–86 (2001).

83. Wright, S. Evolution in Mendelian Populations. *Genetics* **16,** 97–159 (1931).

84. Orr, H. a & Turelli, M. The evolution of postzygotic isolation: accumulating Dobzhansky-Muller incompatibilities. *Evolution* **55,** 1085–94 (2001).

85. Breen, M. S., Kemena, C., Vlasov, P. K., Notredame, C. & Kondrashov, F. a. Epistasis as the primary factor in molecular evolution. *Nature* **490,** 535–538 (2012).

86. Kondrashov, A., Sunyaev, S. & Kondrashov, F. Dobzhansky–Muller incompatibilities in protein evolution. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 14878–14883 (2002).

87. Lappalainen, T., Montgomery, S. B., Nica, A. C. & Dermitzakis, E. T. Epistatic selection between coding and regulatory variation in human evolution and disease. *Am J Hum Genet* **89,** 459–463 (2011).

88. Alvarez-Castro, J. M., Le Rouzic, A., Carlborg, O., Álvarez-Castro, J. M. & Carlborg, Ö. How to perform meaningful estimates of genetic effects. *PLoS Genet.* **4,** e1000062 (2008).

89. Haig, D. Does heritability hide in epistasis between linked SNPs? *Eur. J. Hum. Genet.* **19,** 123 (2011).

90. 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).

91. 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).

92. 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).

93. 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).

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

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

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

97. Wray, N. R. *et al.* Pitfalls of predicting complex traits from SNPs. *Nat Rev Genet* **14,** 507–515 (2013).

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

99. 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).

100. 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).

101. 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).

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

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

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

105. 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

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

107. Greene, C. S. *et al.* Enabling personal genomics with an explicit test of epistasis. *Pac. Symp. Biocomput.* 327–336 (2010). at <http://www.ncbi.nlm.nih.gov/pubmed/19908385>

108. Yang, Y. *et al.* Clinical Whole-Exome Sequencing for the Diagnosis of Mendelian Disorders. *N. Engl. J. Med.* (2013). doi:10.1056/NEJMoa1306555

109. Dorschner, M. O. *et al.* Actionable, pathogenic incidental findings in 1,000 participants’ exomes. *Am. J. Hum. Genet.* **93,** 631–40 (2013).

110. Visscher, P. M. & Gibson, G. What if we had whole-genome sequence data for millions of individuals? *Genome Med.* **5,** 80 (2013).

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

112. Marchini, J., Donnelly, P. & Cardon, L. R. Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nat. Genet.* **37,** 413–417 (2005).

113. Houle, D., Pélabon, C., Wagner, G. & Hansen, T. Measurement and meaning in biology. *Q. Rev. Biol.* **86,** 3–34 (2011).