# Witty, yet serious title to go here

Table of Contents

Witty, yet serious title to go here 1

Abstract 1

Introduction 2

Methods for detecting epistasis 2

What has been found in human complex traits? 3

Lupus 3

Rheumatoid arthritis 3

Type 2 diabetes 3

Alzheimer's Disease 3

Ankylosing spondylitis 3

Psoriasis 3

Gene expression 4

WTCCC 4

Breast cancer 4

Coronary artery disease 4

Systolic blood pressure 4

Bipolar 4

To which scientific question(s) is epistasis the answer? 4

The missing heritability? 4

Elucidating biological mechanisms? 5

Evolutionary persistence of deleterious mutations? 6

Evolutionary impact of complex traits? 6

Improvement in genetic prediction? 6

Discussion 6

Box 1: Why is epistasis theoretically difficult to detect? 7

Box 2: What constitutes a significant epistatic interaction? 8

Glossary 8

References 8

## 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 interactions 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 that one can use to estimate the proportion of the phenotypic variance of a trait that is attributed to additive genetic (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/diseases1-7. The wide range of methods spans from conventional regression-based methods to nature-inspired algorithms (FIG 1). Most methods use SNP-based tests for pair-wise 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 interactions becomes a more routine exercise because the computational barrier has been greatly reduced8-12. Here we try to provide an overview of the many method papers published during this period in an intuitive way to help understand the big picture of the developments.

## What has been found in human complex traits?

The literature is

Mostly very very modest, even when people are looking for it

- Hypothesis-Based Analysis of Gene-Gene Interactions and Risk of Myocardial Infarction

- Shows that narrowing the search for epistasis to amongst marginal effects is not improving matters

- upper bound on epistasis

### Lupus

- Castillejo-López et al. BANK1 and BLK have marginal effects, epistasis is shown and the genes physically interact

### Rheumatoid arthritis

- Genin et al. Same interaction.

### Type 2 diabetes

### Alzheimer's Disease

### Ankylosing spondylitis

### Psoriasis

### Gene expression

### WTCCC

- Wan (BOOST)

- Lippert

### Breast cancer

MDR approach finds 4 way interaction

### Coronary artery disease

- used for prediction but training and validation both within the same sample

### Systolic blood pressure

- Functional epistatic interaction between rs6046G>A in F7 and rs5355C>T in SELE modifies systolic blood pressure levels.

http://www.ncbi.nlm.nih.gov/pubmed/22815813

### Bipolar

Prabhu et al - close to replication but not quite

## 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.2 Such instances might arise if the trait has a large mutational target size and is under direct selection.3 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.

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.4 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 studies provided.5 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,6 and this has indeed been demonstrated . 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, but h2 estimates for BMI are much higher from twin studies than from family studies, suggesting that height is probably mostly influenced by additive effects and non-additive effects may play an important role in BMI.

### Elucidating biological mechanisms?

Suppose that one has identified two polymorphisms, from independent regions of the genome, that exhibit a statistical interaction. What biological information can this tell us?

functions of mutations

relationships between independent genomic features

construction of pathways

genetic redundancy

Are there examples of interactions leading to understanding?

Are there examples where this hasn't improved understanding?

Epistasis between COMT and MTHFR in Maternal-Fetal Dyads Increases Risk for Preeclampsia

- examples of epistatic interaction being hypothesised because of known biological function

- (see also LOAD example)

Epistasis: Obstacle or Advantage for Mapping Complex Traits? & Hemani 2013

- Suggest that searching for epistatic effects will find SNPs that have no marginal associations

### Evolutionary persistence of deleterious mutations?

### Evolutionary impact of complex traits?

Much has

### Improvement in genetic prediction?

- Gustavo and Dan Gianola

## Discussion

Scale effects, haplotype effects, linked loci, liability vs observed scale

Should we be looking for epistasis?

For:

1. They may explain missing variation

2. They may indicate the evolutionary persistence of additive effects

3. It seems to be widespread in artificial genetic variation

4. Knowing how independent regions interact statistically could shed light on biological mechanisms

Against:

1. The additive model is simple and effective

2. There are no non-additive models in animal breeding but there is extremely good progress in morphological selection

3. Detection of epistatic effects is very hard, and additive models are already underpowered

4. We expect less variance to be contributed by non-additive effects based on Taylor series expansions, narrow sense heritability estimates, etc.

============================================================================

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

Higher order interactions

Curse of dimensionality / significance thresholds

Sample size

Replication - Sample LD is larger for r^8 than it is for r^2, ascertainment of high r in discovery leads to low replication in replication - need to use imputed data.

The genetic variance of complex traits is typically comprised of very many polymorphisms, each with a very small effect. Herein lies the most parsimonious explanation for the problem of the missing heritability, and informs that association studies need to increase in size dramatically in order to detect more variants. Supposing that non-additive variance is similarly comprised of numerous small effects, 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 r^2 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 decrease in dominance variance with LD r^2 is quadratic, 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 additive x additive variance is linearly dependent upon sufficient LD at two independent positions (reduces quadratically with LD r^2), and dominance x dominance variance is quadratically dependent at both positions (reduces to the fourth power with LD r^2). 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 larger sample sizes or more dense genotyping (or sequencing) to identify non-additive effects at the same power as additive effects.

- Curse of dimensionality / significance thresholds

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

Replication

Higher order interactions

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

Mustn't be tolerant of false positives just because detection is hard

Replication is necessary

## Glossary

Complex traits

Mutational target size

Additive genetic variance

Marginal effects

## References

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

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

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

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

5. Zuk, O., Hechter, E., Sunyaev, S. R. & Lander, E. S. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences* (2012). doi:10.1073/pnas.1119675109

6. Evans, D. M., Gillespie, N. a & Martin, N. G. Biometrical genetics. *Biological psychology* **61,** 33–51 (2002).