
Andrew G Clark
Cornell University
Missing Heritability II
1-2 May 2018 – Silver Spring, MD

# Objectives

- There can be pervasive epistasis and yet additive models fit the data well.
- Gene-gene interaction can be strong and still generate little epistatic variance.
- The Infinitesimal model – is clearly silly, but it still has great utility.
- Infinitesimal epistatic model does not improve fit.
- Individual prediction that accommodates gene-gene interaction and GxE may yet be useful.

# Epistasis in Manolio *et al*. (2009)

- "Narrow-sense heritability estimates in humans can be inflated if family resemblance is influenced by non-additive genetic effects (dominance and epistasis, or gene–gene interaction), shared familial environments, and by correlations or interactions among genotypes and environment."
- "Box 2: To investigate missing heritability using family studies, the following measures are required: ….(8) Identify gene–gene interactions by positive correlation between family-specific logs odds ratio (lod) scores or evidence of linkage disequilibrium among unlinked loci."
- "Box 3: The following steps can be used to make the most of existing and future GWAS: …. (7) Investigate gene–gene interactions, including dominance and epistasis."

# Gene-gene interaction without epistatic variance

Consider a multi-locus model for a quantitative trait:

| Genotype | Genotypic values without dominance |
|---|---|
| $a_1a_1\ a_2a_2\ a_3a_3$ | $0$ |
| $a_1a_1\ a_2a_2\ A_3a_3$ | $a_3$ |
| $a_1a_1\ a_2a_2\ A_3A_3$ | $2a_3$ |
| $a_1a_1\ A_2a_2\ A_3a_3$ | $a_2 + a_3 + [aa]_{23}$ |
| $a_1a_1\ A_2a_2\ A_3A_3$ | $a_2 + 2a_3 + 2[aa]_{23}$ |
| $a_1a_1\ A_2A_2\ A_3A_3$ | $2a_2 + 2a_3 + 4[aa]_{23}$ |
| $A_1a_1\ A_2a_2\ A_3a_3$ | $a_1 + a_2 + a_3 + [aa]_{12} + [aa]_{13} + [aa]_{23} + [aaa]_{123}$ |
| $A_1a_1\ A_2a_2\ A_3A_3$ | $a_1 + a_2 + 2a_3 + [aa]_{12} + 2[aa]_{13} + 2[aa]_{23} + 2[aaa]_{123}$ |
| $A_1a_1\ A_2A_2\ A_3A_3$ | $a_1 + 2a_2 + 2a_3 + 2[aa]_{12} + 2[aa]_{13} + 4[aa]_{23} + 4[aaa]_{123}$ |
| $A_1A_1\ A_2A_2\ A_3A_3$ | $2a_1 + 2a_2 + 2a_3 + 4[aa]_{12} + 4[aa]_{13} + 4[aa]_{23} + 8[aaa]_{123}$ |

Mäki-Tanila & Hill 2014 *Genetics* 198:355

# Additive variance has epistatic terms

The average effect of an allele can be written as:

$$\alpha_i = \frac{1}{2}\frac{\partial \mu}{\partial p_i} = a_i + 2\sum_{j \neq i} p_j [aa]_{ij} + 4\sum_{j \neq i}\sum_{k \neq i, k > j} p_j p_k [aaa]_{ijk}$$

And the additive variance is the sum of the squared effects weighted by frequencies:

$$V_A = \sum_i 2p_i(1-p_i)\left(\frac{1}{2}\partial\mu/\partial p_i\right)^2 = \sum_i H_i \alpha_i^2,$$

Kojima 1959 *Genetics* 45: 984
Mäki-Tanila & Hill 2014 *Genetics* 198: 355

# Gene-Gene interactions contribute to $V_A$

$$V_A = \sum_i H_i \Big( a_i^2 + 4p_j a_i [aa]_{ij} + 4p_k a_i [aa]_{ik} + 4p_j^2 [aa]_{ij}^2$$
$$+4p_k^2 [aa]_{ik}^2 + 8p_j p_k [aa]_{ij}[aa]_{ik} + 8p_j p_k a_i [aaa]_{ijk}$$
$$+16p_j^2 p_k [aa]_{ij}[aaa]_{ijk} + 16p_j p_k^2 [aa]_{ik}[aaa]_{ijk}$$
$$+16\, p_j^2 p_k^2 [aaa]_{ijk}^2 \Big), \quad \text{where } i \neq j \neq k,$$

**Punchline: A substantial portion of variation that is caused by epistatic interaction ends up in the additive variance (contributing to heritability).**

"a rather mundane theoretical finding"

Mäki-Tanila & Hill 2014 *Genetics* 198:355

# Fisher's Infinitesimal Model

- Very large number of unlinked loci, each with very small effect.
- Assume HW and linkage equililbrium.
- Phenotype is obtained as the sum of allelic effects, and is normally distributed.

Barton, Etheridge & Véber 2017 *Theor Pop Biol*

# Fisher's Infinitesimal Model: results

- Variance of offspring does not depend on trait values of the parents.
- Selection produces negligible change in allele frequency (or variance).
- The Breeder's Equation ($R = h^2 s$) follows.
- The model can accommodate epistasis.
- Consequences of stabilizing selection, inbreeding and assortative mating are easily derived.

Barton, Etheridge & Véber 2017 *Theor Pop Biol*

# Infinitesimal model of epistasis

- Suppose each pairwise interaction is small.
- Only very few are genome-wide significant.
- But there are $O(n^2)$ such pairwise interactions.
- So in aggregate, perhaps they can contribute to genotype-phenotype association.
- Can we fit this model, just as Yang *et al*. (2010) did for additive effects?
- Barton, Etheridge & Véber (2017) – deviation from infinitesimal model is $1/\sqrt{M}$, where $M$ is the number of loci.

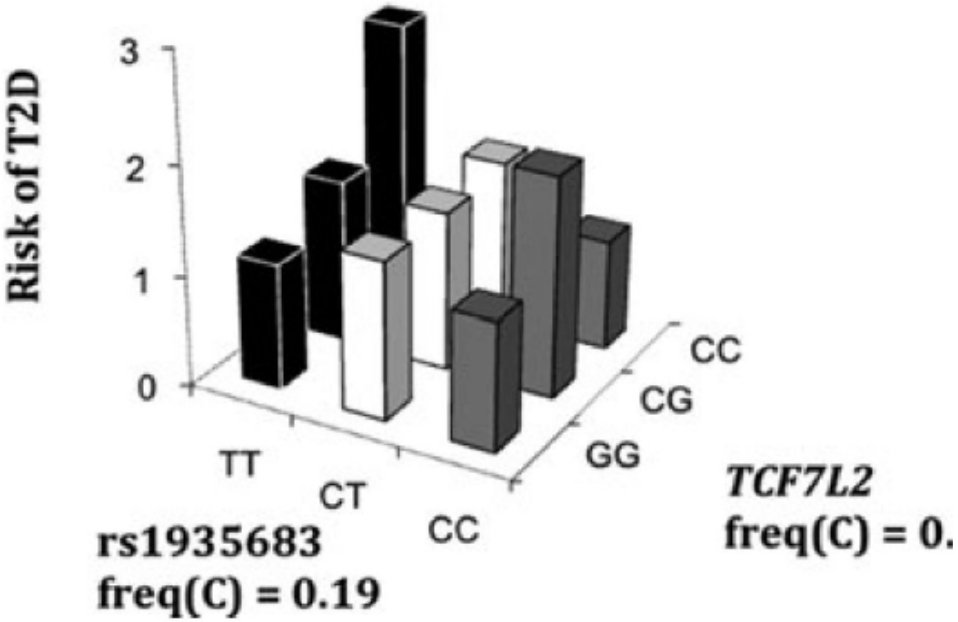
## Extending the infinitesimal model to data on the whole genome…

"… regression of a trait on sequence can significantly improve predictions of breeding value, even when individual loci cannot be identified: this is the basis of ''genomic selection'' (Meuwissen et al., 2013).

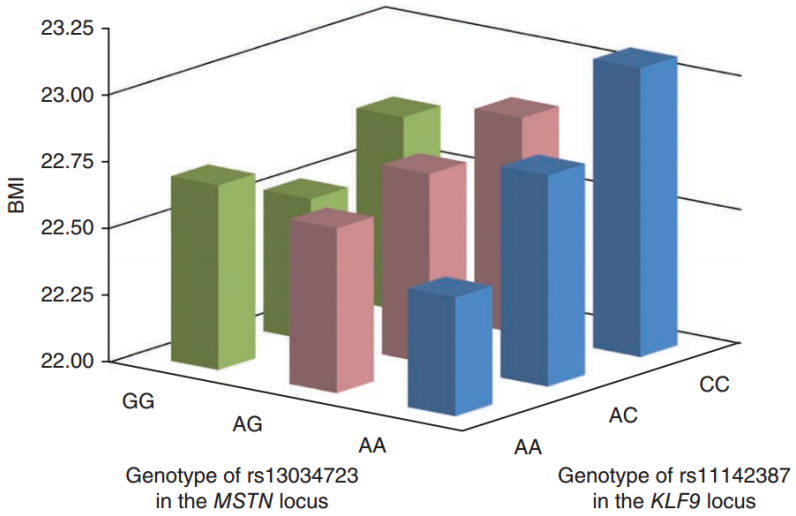
It may be that natural selection is in just the same position as a breeder: selection may change the mean rapidly and predictably, even when the probability distribution of any particular allele frequency is hardly perturbed."

# Pairwise epistasis from GWAS

## Type 2 Diabetes




## Body Mass Index

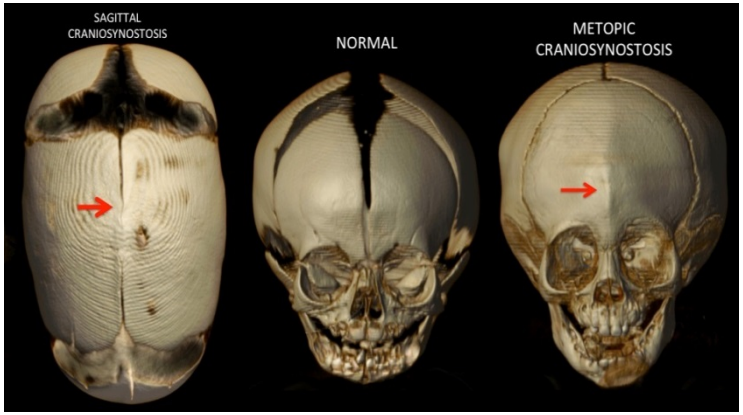


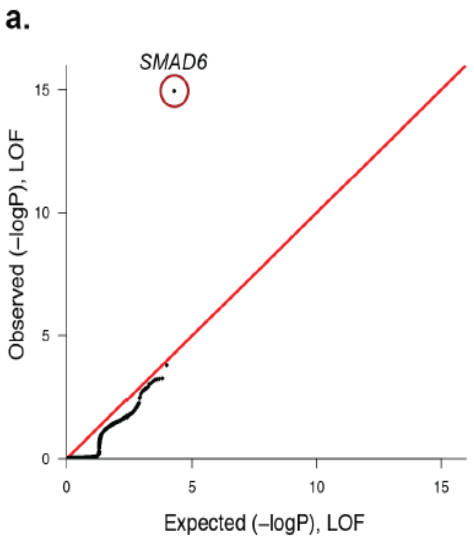

Bell *et al.* 2011 *Ann Hum Genet*

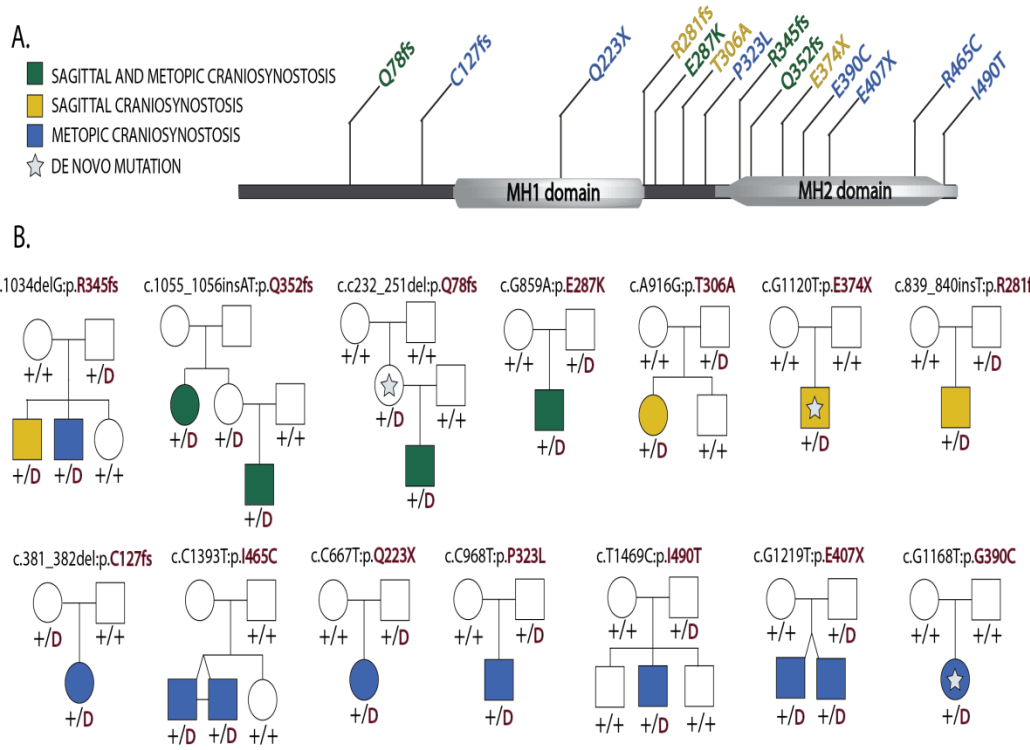
Okada *et al.* 2012 *Nature Genet*

# Pairwise epistasis from Mendelian disorders

## Non-syndromic midline craniosynostosis




Phenotypically normal carrier parents of the dominant Smad6 mutation all have common variants in BMP2.

Timberlake *et al*. 2016 *eLife*

# So does epistasis matter in evolution?

- Genes exist in networks and epistasis at the molecular level is pervasive.
- And yet, the infinitesimal model fits the data.
- Paixão & Barton (2016) compare a purely additive model to one with epistasis.
- If selection is weak ($Ns < 1$), drift dominates and variance components are unchanged (infinitesimal model of nonadditive effects).
- If selection is strong ($Ns > 1$), allele frequencies change, and the genotype-phenotype map matters more than variance components.

Paixão & Barton 2016 *PNAS*

# Heritability vs. Individual prediction

- Genomic prediction is for Breeding Value (mean phenotype of many offspring).
- GxG and GxE may perturb each individual from this expectation.
- The ideal individual prediction could accommodate GxG and GxE, if only we could estimate them …

# Machine Learning to the rescue ??



2016

## Genomic Selection with Deep Neural Network

Riley Mitchell Mcdowell
*Iowa State University*

## Application of Deep Learning in Genomic Selection

Yang Liu
Informatics Institute,
University of Missouri, Columbia
Columbia, MO USA
Ylmk2@mail.missouri.edu

Duolin Wang
Christopher S. Bond Life Sciences
University of Missouri, Columbia
Columbia, MO USA
wangdu@missouri.edu

## An Evaluation of Machine-learning for Predicting Phenotype: studies in yeast and wheat

Nastasiya F. Grinberg[1] Ross D. King[2]

February 2, 2017

1. Department of Medicine, University of Cambridge Cambridge CB2 0SP
2. School of Computer Science, University of Manch 9PL



## DeepGS: Predicting phenotypes from genotypes using Deep Learning

Wenlong Ma[1,†], Zhixu Qiu[1,†], Jie Song[1], Qian Cheng[1,2], Chuang Ma[1,2,*]

[1]State Key Laboratory of Crop Stress Biology for Arid Areas, Center of Bioinformatics, College of Life Sciences, Northwest A&F University, Yangling, Shaanxi 712100, China, [2]Key Laboratory of Biology and Genetics Improvement of Maize in Arid Area of Northwest Region, Ministry of Agriculture, Northwest A&F University, Shaanxi, Yangling 712100, China.

# Reasons for missing epistasis

- Markers are in imperfect LD with causal variants.
- Rapid population growth → rare alleles.
- Curse of dimensionality (power).
- Small effect size (power).
- Embedding in higher dimension gene-gene interactions (epistasis appears as additive variance)
- Embedding in gene x environment interactions.
- Population substructure (heterogeneous embedding).