Genome-wide strategies for detecting multiple loci that influence complex diseases

Jonathan Marchini, Peter Donnelly, Lon R Cardon

Presented by Jeff Kilpatrick

Introduction

Introduction

- Genetic epidemiologists have unprecedented mountains of data

Introduction

- Genetic epidemiologists have unprecedented mountains of dat

Introduction

- Genetic epidemiologists have unprecedented mountains of data
- Large collections of human data now available

Introduction

- Genetic epidemiologists have unprecedented mountains of data
- Large collections of human data now available
- Massively parallel genotyping can produce data for over a million genetic markers per person -- fast

Introduction

Introduction

- Great! So here's the plan:

Introduction

- Great! So here's the plan:

  1. Evaluate each marker for association with disease

Introduction

- Great! So here's the plan:

  1. Evaluate each marker for association with disease

  2. Compile list of genes near significant markers

Introduction

- Great! So here's the plan:

  1. Evaluate each marker for association with disease

  2. Compile list of genes near significant markers

  3. Publish in Nature

Introduction

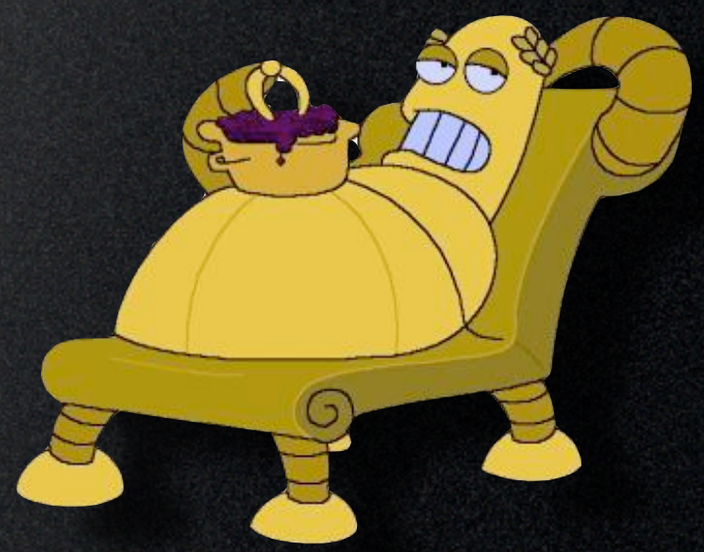
- Great! So here's the plan:

  1. Evaluate each marker for association with disease

  2. Compile list of genes near significant markers

  3. Publish in Nature

  4. Grow fat and wealthy with a supermodel spouse



Introduction

Introduction

- Wake up! The reality of genotype-phenotype association Hell:

Introduction

- Wake up! The reality of genotype-phenotype association Hell:
- Evidence suggests interactions contribute broadly to complex traits

Introduction

- Wake up! The reality of genotype-phenotype association Hell:
  - Evidence suggests interactions contribute broadly to complex traits
  - Frequency distribution of marker variants affects their statistical power

Introduction

- This paper explores two questions
  - Is there hope for consistently detecting such effects?
  - How do we design and analyze genome-wide association studies?

The Plan Today

- Interaction models
- Analysis strategies
- Power analysis
- Loose ends

- Interaction models
- Analysis strategies
- Power analysis
- Loose ends

Interaction Models

Interaction Models

- Model: a mathematical description of how genes confer risk

Interaction Models

- Model: a mathematical description of how genes confer risk
- Example: “exactly two disease variants from two susceptibility loci are required”

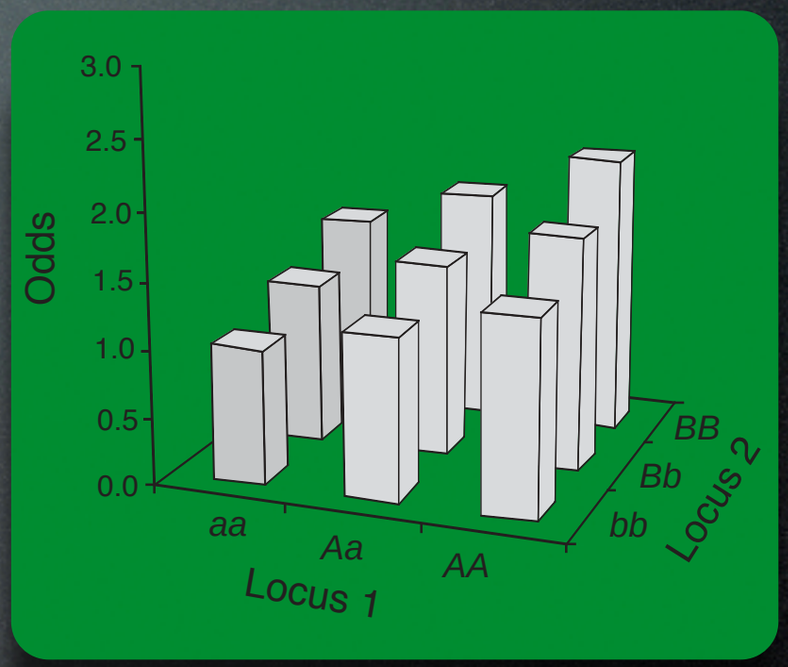
Interaction Models

- The example:

	AA	Aa	aa
BB			✓
Bb		✓	
bb	✓		

Interaction Models

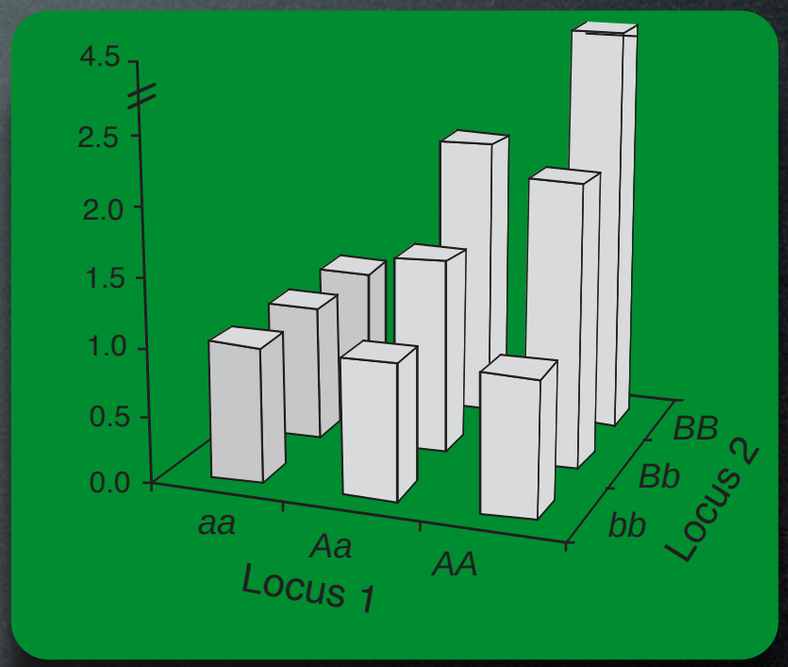
- Adding a disease variant at either marker multiplicatively increases risk
- Loci do not interact



Model 1: multiplicative within and between loci

Interaction Models

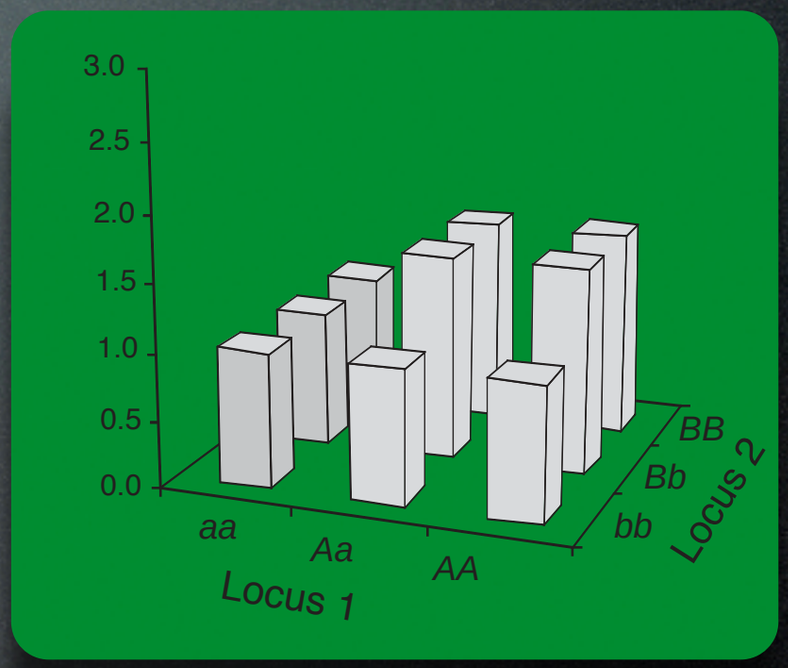
- Neither locus alone is sufficient
- Multiple risk alleles from different loci increase risk linearly



Model 2: two-locus interaction multiplicative effects

Interaction Models

- Neither locus alone is sufficient
- Presence of risk variants from both markers increases elevates risk to constant level



Model 3: two-locus interaction threshold effects

- Interaction models
- Analysis strategies
- Power analysis
- Loose ends

Analysis Strategies

Analysis Strategies

- Outside our dream world, we have to be selective in the tests we conduct

Analysis Strategies

- Outside our dream world, we have to be selective in the tests we conduct
- Tests cost time. Time is money.

Analysis Strategies

- Outside our dream world, we have to be selective in the tests we conduct
- Tests cost time. Time is money.
- Tests cost significance

Analysis Strategies

Analysis Strategies

- Strategy I -- “Dreamland”

Analysis Strategies

- Strategy I -- "Dreamland"
  - Perform locus-by-locus search

Analysis Strategies

- Strategy I -- "Dreamland"
  - Perform locus-by-locus search
    - For n markers, n tests are required

Analysis Strategies

- Strategy I -- "Dreamland"
  - Perform locus-by-locus search
  - For n markers, n tests are required
  - Has a snowball's chance to discover interactions

Analysis Strategies

- Strategy I -- "Dreamland"
  - Perform locus-by-locus search
    - For n markers, n tests are required
    - Has a snowball's chance to discover interactions

Analysis Strategies

Analysis Strategies

- Strategy II -- "Styx"

Analysis Strategies

- Strategy II -- "Styx"
  - Test all pairs of loci

Analysis Strategies

- Strategy II -- "Styx"
  - Test all pairs of loci
  - Requires n^2 tests

Analysis Strategies

- Strategy II -- "Styx"
  - Test all pairs of loci
  - Requires n^2 tests
  - Will discover all pairwise interactions, assuming their effects survive correction for multiple tests

Analysis Strategies

Analysis Strategies

- Strategy III -- "The Compromise"

Analysis Strategies

- Strategy III -- "The Compromise"
  - Search for mildly associated loci

Analysis Strategies

- Strategy III -- "The Compromise"
  - Search for mildly associated loci
  - All pairs of selected loci are tested

- Interaction models
- Analysis strategies
- Power analysis
- Loose ends

Power Analysis

Power Analysis

- Simulated genotypes generated at two loci under each model

Power Analysis

- Simulated genotypes generated at two loci under each model
- Calculations assume L = 300,000 markers, with two (unobserved) causative loci

Power Analysis

- Simulated genotypes generated at two loci under each model
- Calculations assume L = 300,000 markers, with two (unobserved) causative loci
- Bonferroni correction applied

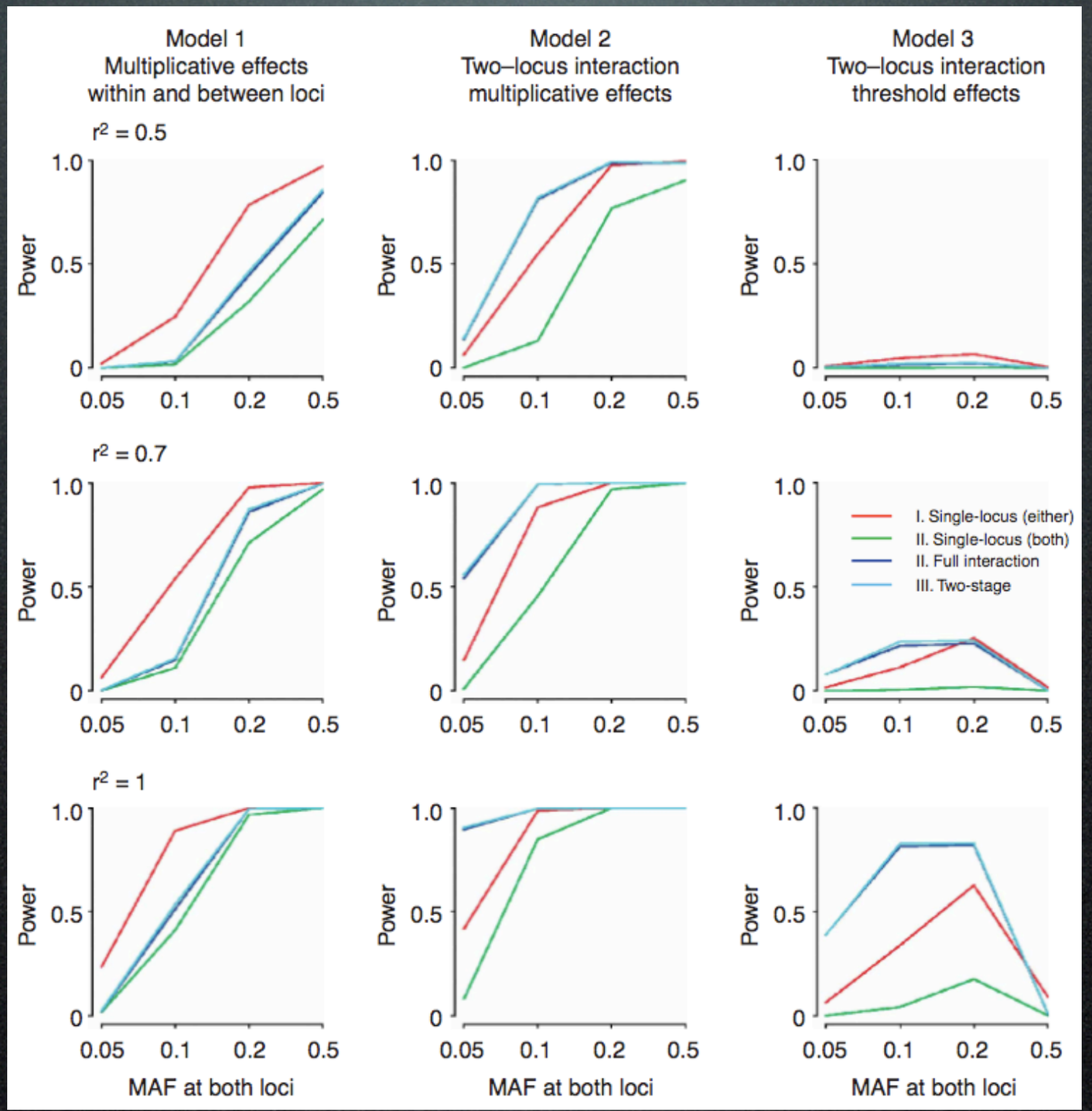
Power Analysis

Distance to disease locus

High

Medium

Low



Dreamland (either locus)

Dreamland (both loci)

Styx (both loci)

The Compromise (both loci)

Power Analysis

Power Analysis

- Interaction-based searches perform well, in spite of harsh correction

Power Analysis

- Interaction-based searches perform well, in spite of harsh correction
  - Except when recovering one marker under Model 1

Power Analysis

- Interaction-based searches perform well, in spite of harsh correction
  - Except when recovering one marker under Model 1
- Power strongly correlated with minor allele frequency and LD

Power Analysis

Power Analysis

- All three strategies are computationally feasible

Power Analysis

- All three strategies are computationally feasible
- Styx approach took 33 hours on ten nodes with 300,000 markers and 2,000 subjects

- Interaction models
- Analysis strategies
- Power analysis
- Loose ends

Loose Ends

Loose Ends

- Power analysis suggests reasons for failure to replicate

Loose Ends

- Power analysis suggests reasons for failure to replicate
- Presence of locus interaction

Loose Ends

- Power analysis suggests reasons for failure to replicate
- Presence of locus interaction
- Different allele frequencies between initial and follow-up cohorts

Loose Ends

Loose Ends

- This study understates usefulness of interaction searches

Loose Ends

- This study understates usefulness of interaction searches
- Bonferroni is conservative

Loose Ends

- This study understates usefulness of interaction searches
- Bonferroni is conservative
- Permutation testing would be more accurate

Conclusions

Conclusions

- All non-exhaustive interaction searches may miss some effects

Conclusions

- All non-exhaustive interaction searches may miss some effects
- Complete enumeration is too expensive for higher order effects

Conclusions

- All non-exhaustive interaction searches may miss some effects
- Complete enumeration is too expensive for higher order effects
- The Compromise provides the best of both worlds in most studies

Questions