Analyzing SNP Data

• Study Design
• SNPs vs Haplotypes
• Regression Analysis
• Population Structure
• Multiple Testing
• Whole Genome Analysis

Heritability

• Is your favorite phenotype genetic?
• Heritability ($h^2$) is the proportion of variance attributed to genetic factors
  – $h^2 < 100\%$: ABO Blood type, CF
  – $h^2 > 80\%$: Height, BMI, Autism
  – $h^2 50-80\%$: Smoking, Hypertension, Lipids
  – $h^2 20-50\%$: Marriage, Suicide, Religiousness
  – $h^2 \sim 0\%$: ??

Prior Hypotheses

• There will always be too much data
• There will (almost) always be priors
  – Favored SNPs
  – Favored Genes
• Make sure you’ve stated your priors (if any) explicitly BEFORE you look at the data

Study Design

• Heritability
• Prior hypotheses
• Target phenotype(s)
• Power
• Ethnicity
• Replication
Target Phenotypes

- LDL
- Diet
- MI
- IL6
- Acute Illness

Carlson et al., Nature v. 429 p. 446

Statistical Power

- Null hypothesis: all alleles are equal risk
- Given that a risk allele exists, how likely is a study to reject the null?
- Are you ready to genotype?

Genetic Relative Risk

\[
RR = \frac{p(Disease|Allele1)}{p(Disease|Allele2)} = \frac{p_{1D}}{p_{2D}} = \frac{p_{1D} + p_{1U}}{p_{2D} + p_{2U}}
\]

Power Analysis

- Statistical significance
  - Significance = p(false positive)
  - Traditional threshold 5%
- Statistical power
  - Power = 1- p(false negative)
  - Traditional threshold 80%
- Traditional thresholds balance confidence in results against reasonable sample size

Small sample: 50% Power

Maximizing Power

- Effect size
  - Larger relative risk = greater difference between means
- Sample size
  - Larger sample = smaller SEM
- Measurement error
  - Less error = smaller SEM
Large sample: 97.5% Power

Risk Allele Example
10% Population Frequency

- Homozygous
  Relative Risk = 4
- Homozygous
  Relative Risk = 2
- Multiplicative Risk Model
  - Het RR = 2
- Multiplicative Risk Model
  - Het RR = 1.4
- Case Freq
  - 18.2%
- Case Freq
  - 13.6%
- Control Freq
  - 9.9%
- Control Freq
  - 9.96%

Power to Detect RR=2
N Cases, N Controls

Power to Detect RR=2
N Cases, N Controls

Power to Detect RR=2
N Cases, N Controls

Power to Detect RR=2
N Cases, N Controls
Power to Detect SNP Risk
200 Cases, 200 Controls

![Power Analysis Summary]

- For common disease, relative risk of common alleles is probably less than 4
- Maximize number of samples for maximal power
- For RR < 4, measurement error of more than 1% can significantly decrease power, even in large samples

SNP Selection for Association Studies

Direct:
Catalog and test all functional variants for association

Indirect:
Use dense SNP map and select based on LD

Parameters for SNP Selection

- Allele Frequency
- Putative Function (cSNPs)
- Genomic Context (Unique vs. Repeat)
- Patterns of Linkage Disequilibrium

Focus on Common Variants - Haplotype Patterns

All Gene SNPs
SNPs > 10% MAF

Why Common Variants?

- Rare alleles with large effect (RR > 4) should already be identified from linkage studies
- Association studies have low power to detect rare alleles with small effect (RR < 4)
- Rare alleles with small effect are not important, unless there are a lot of them
- Theory suggests that it is unlikely that many rare alleles with small effect exist (Reich and Lander 2001).
Ethnicity

All Gene SNPs

SNPs > 10% MAF

African American

European American

Replication

- You WILL be asked to replicate
- Statistical replication
  - Split your sample
  - Arrange for replication in another study
  - Multiple measurements in same study
- Functional replication

Multiple Measurements:

CRP in CARDIA

Haplotype vs tagSNPs

High CRP Haplotype

- 5 SNPs specific to high CRP haplotype

Functional Replication

- Statistical replication is not always possible
- Association may imply mechanism
- Test for mechanism at the bench
  - Is predicted effect in the right direction?
  - Dissect haplotype effects to define functional SNPs

Carlson et al. AJHG v77 p64

Haplo.glm: Lake et al. Hum Hered v. 55 p. 56
CRP Evolutionary Conservation

- TATA box: 1697
- Transcript start: 1741
- CRP Promoter region (bp 1444-1650) >75% conserved in mouse

Low CRP Associated with H1-4

- USF1 (Upstream Stimulating Factor)
  - Polymorphism at 1440 alters USF1 binding site
    1420     1430     1440
    H1-4 gcagctacCACGTGcacccagatggcCACTCGtt
    H7-8 gcagctacCACGTGcacccagatggcCACTAGtt
    H5-6 gcagctacCACGTGcacccagatggcCACTTGGtt

High CRP Associated with H6

- USF1 (Upstream Stimulating Factor)
  - Polymorphism at 1421 alters another USF1 binding site
    1420 1430 1440
    H1-4 gcagctacCACGTGcacccagatggcCACTCGtt
    H7-8 gcagctacCACGTGcacccagatggcCACTAGtt
    H5 gcagctacCACGTGcacccagatggcCACTTGGtt
    H6 gcagctacCACATGcacccagatggcCACTTGGtt

CRP Promoter Luciferase Assay

- Carlson et al, AJHG v77 p64

CRP Gel Shift Assay

- Szalai et al, J Mol Med v83 p440

Study Design Summary

- State your priors
- Know your phenotypes
- Estimate your power
- Pay attention to ethnicity
- Set up replication ASAP
- Replication can be functional
Data Analysis

• Study Design
• SNPs vs Haplotypes
• Regression Analysis
• Population Structure
• Multiple Testing
• Whole Genome Analysis

SNPs or Haplotypes

• There is no right answer: explore both
• The only thing that matters is the correlation between the assayed variable and the causal variable
• Sometimes the best assayed variable is a SNP, sometimes a haplotype

Example: APOE

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Population (%)</th>
<th>AD (%)</th>
<th>#Population</th>
<th>1AD (%)</th>
<th>Risk (%)</th>
<th>If all US</th>
</tr>
</thead>
<tbody>
<tr>
<td>ε2/ε2</td>
<td>1</td>
<td>0.1</td>
<td>0.096M</td>
<td>0.27</td>
<td>0.05</td>
<td>0.08</td>
</tr>
<tr>
<td>ε2/ε3</td>
<td>12</td>
<td>4</td>
<td>0.18M</td>
<td>0.27</td>
<td>0.35</td>
<td>0.38</td>
</tr>
<tr>
<td>ε2/ε4</td>
<td>69</td>
<td>35</td>
<td>0.06M</td>
<td>1.9M</td>
<td>1.54</td>
<td>1.8M</td>
</tr>
<tr>
<td>ε3/ε3</td>
<td>16</td>
<td>9.8M</td>
<td>1.7M</td>
<td>14.8M</td>
<td>18.8M</td>
<td>20.3M</td>
</tr>
<tr>
<td>ε3/ε4</td>
<td>2</td>
<td>0.95M</td>
<td>0.06M</td>
<td>0.67</td>
<td>30.7M</td>
<td>0.23</td>
</tr>
</tbody>
</table>

Please note that ε5/4 subjects are not included in table.


b Assuming 4 million individuals have AD.

c Data from [3,46,49].

Raber et al, Neurobiology of Aging, v25 p641

Example: APOE

• Small gene (<6kb)
• 7 SNPs with MAF > 5%
• APOE ε2/3/4
  – Alzheimer’s associated
  – ε2 = 4075
  – ε4 = 3937

Example: APOE

• Only three meaningful categories of haplotype
• No single SNP is adequate
Example: APOE

• SNP analysis:
  – 7 SNPs
  – 7 tests with 1 d.f.

• Haplotype analysis
  – 13 haplotypes
  – 1 test with 12 d.f.

Best marker is a haplotype of only the right two SNPs:
3937 and 4075

Building Up

• Test each SNP for main effect

• Test SNPs with main effects for interactions

Paring Down

• Test all haplotypes for effects

• Merge related haplotypes with similar effect

Paring Down

• Test all haplotypes for effects

Data Analysis

• Study Design
• SNPs vs Haplotypes
• Regression Analysis
• Population Structure
• Multiple Testing
• Whole Genome Analysis
Exploring Candidate Genes:
Regression Analysis

• Given
  – Height as “target” or “dependent” variable
  – Sex as “explanatory” or “independent” variable
• Fit regression model
  \[ \text{height} = \beta \times \text{sex} + \varepsilon \]

Regression Analysis

• Given
  – Quantitative “target” or “dependent” variable \( y \)
  – Quantitative or binary “explanatory” or “independent” variables \( x_i \)
• Fit regression model
  \[ y = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_i x_i + \varepsilon \]

Regression Analysis

• Works best for normal \( y \) and \( x \)
• Fit regression model
  \[ y = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_i x_i + \varepsilon \]
• Estimate errors on \( \beta \)'s
• Use t-statistic to evaluate significance of \( \beta \)'s
• Use F-statistic to evaluate model overall

Call:
\[ \text{lm(formula = data}\_\text{TARGET} \sim (\text{data}\_\text{CURR}\_\text{AGE} + \text{data}\_\text{CIGNOW} + \text{data}\_\text{PACKYRS} + \text{data}\_\text{SNP1} + \text{data}\_\text{SNP2} + \text{data}\_\text{SNP3} + \text{data}\_\text{SNP4})) \]

Residuals:
- Min       1Q   Median     3Q      Max
  -123.425   -25.794   -3.125    23.629  120.046

Coefficients:  Estimate Std. Error t value Pr(>|t|)
(Intercept)           139.52703   13.80820  10.105  < 2e-16 ***
data\_\text{CURR}\_\text{AGE}          -0.04844    0.18492  -0.262  0.79345
data\_\text{CIGNOW}           -10.11001    4.06797  -2.485  0.01327 *
data\_\text{PACKYRS}             0.01573    0.05456   0.288  0.77320
data\_\text{SNP1}               8.61749    3.31204   2.602  0.00955 **
data\_\text{SNP2}             -19.71980    2.84816  -6.924 1.35e-11 ***
data\_\text{SNP3}              -9.32590    2.96600  -3.144  0.00176 **
data\_\text{SNP4}              -9.58801    3.05650  -3.137  0.00181 **

---
Signif. codes:  0 ** *** 0.001 ** 0.01 * 0.05 `.` 0.1 ` ' 1

Residual standard error: 36.11 on 503 degrees of freedom
Multiple R-Squared: 0.2551, Adjusted R-squared: 0.2448
F-statistic: 24.61 on 7 and 503 DF,  p-value: < 2.2e-16

Coding Genotypes

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Dominant</th>
<th>Additive</th>
<th>Recessive</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>AG</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>GG</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

• Genotype can be re-coded in any number of ways for regression analysis
• Additive ~ codominant

Fitting Models

• Given two models
  \[ y = \beta_1 x_1 + \varepsilon \]
  \[ y = \beta_1 x_1 + \beta_2 x_2 + \varepsilon \]
  • Which model is better?
  • More parameters will always yield a better fit

Information Criteria
  • Measure of model fit penalized for the number of parameters in model
  • AIC (most common)
  • Akaike’s Info Criterion
  • BIC (more stringent)
  • Bayesian Info Criterion
Tool References

- Haplo.stats (haplotype regression)
- PHASE (case/control haplotype)
- Haplo.view (case/control SNP analysis)
- SNPHAP (haplotype regression?)

Analyzing SNP Data

- Study Design
- SNPs vs Haplotypes
- Regression Analysis
- Population Structure
- Multiple Testing
- Whole Genome Analysis

Population Stratification

- Many diseases have different frequencies in ancestral groups
  - E.g. MS is more frequent in Europeans
- In admixed or stratified populations, markers correlated with ancestry may show spurious associations
  - E.g. Duffy and MS in African Americans

Population Stratification

- Admixture
  - Individuals with ancestry from multiple populations
  - E.g. Hispanic or African American
- Stratification
  - Subpopulations with distinct allele frequencies
  - E.g. Brazil, California
- STRUCTURE software
  - Pritchard et al, Genetics v155 p945

Genomic Controls

- Unlinked anonymous markers not chosen for known allele frequencies
- Allow unbiased estimation of population structure

Genomic Controls

- Warning: 377 microsatellites barely detects European structure
- Within continent resolution probably requires thousands of SNPs
Ancestry Informative Markers (AIMs)

- Markers with known allele frequency differences between ancestral groups
- E.g. Duffy blood group
- Useful in estimating ancestry of admixed individuals
- Only relevant to defined ancestral populations

Admixture mapping

- Type several thousand AIMs
- Search for regions with excess allelic ancestry from a single population
- E.g. MS in AA: Reich et al, Nat Genet v37 p1113

Pop Structure Summary

- For known admixture, use AIMs to estimate ancestry
- For diseases with substantial differences in risk by ethnicity, use admixture mapping
- Detecting cryptic population structure requires hundreds to thousands of genomic controls

Analyzing SNP Data

- Study Design
- SNPs vs Haplotypes
- Regression Analysis
- Population Structure
  - Multiple Testing
  - Whole Genome Analysis

Multiple Testing

<table>
<thead>
<tr>
<th>Study target</th>
<th>Technology</th>
<th>Samples</th>
<th>Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene 10 SNPs</td>
<td>TaqMan</td>
<td>100’s</td>
<td>2</td>
</tr>
<tr>
<td>Pathway 1500 SNPs</td>
<td>Illumina SNPlex</td>
<td>1000’s</td>
<td>2</td>
</tr>
<tr>
<td>Genome 500k SNPs</td>
<td>Affy Illumina</td>
<td>??</td>
<td>??</td>
</tr>
</tbody>
</table>

Multiple Testing

- Practical guidelines
  - Write down your priors
  - Bonferroni
  - FDR
  - Staged Study Design
  - Other approaches - Neural Nets
Bonferroni

- P-values of stats assume a single test
- For multiple tests, adjust significance by multiplying P-value by number of tests
  - Given 10 tests and unadjusted $p = 0.02$
  - $p = 10 \times 0.02 = 0.2$
- Over conservative

Step-Down Bonferroni

- Given N SNPs to analyze
- Order SNPs to analyze
  - Evaluate the most interesting hypotheses first
- For first SNP, do not correct p-value
- For second SNP, adjust for 2 tests
- Etc.

Staged Study Design

- Given 500,000 SNPs
- Bonferroni corrected significance threshold
  - $p = \frac{0.05}{500000} = 10^{-7}$
- Significance in a single study is difficult to achieve

Step-Down Bonferroni

- Study I: Genotype 500k SNPs in 1000 cases/controls
  - Expect 5,000 false positives at $p < 0.01$
- Study II: Genotype best 5000 hits from stage I in additional 1000 cases/controls
  - Expect 50 false positives at $p < 0.01$
- Study III: Genotype best 50 hits in a third set of 1000 cases/controls
  - Expect 0.5 false positives at $p < 0.01$

Joint Analysis

- Skol et al, Nat Genet in press

Post-Hoc Analysis

- Significance
  - Probability of a single observation under $H_0$
- False Discovery Rate
  - Proportion of observed results inconsistent with $H_0$
FDR Example

• Assume 10 tests
• 5 with uncorrected \( p = 0.05 \)
• No single significant result
• More than 5% below 5%
• At least one of the five is probably real, but we can’t say which

Multiple Testing Summary

• Bonferroni can be useful, but overly conservative
• FDR can be more helpful
• Staged study designs don’t improve power, but can be economically advantageous

Analyzing SNP Data

• Study Design
• SNPs vs Haplotypes
• Regression Analysis
• Population Structure
• Multiple Testing
• Whole Genome Analysis

SNP Selection

• cSNPs (~20-25k common genome wide)
• tagSNPs
  – 500k random = 300k selected
  – Probably adequate in European
  – Possibly adequate in Asian
  – More needed for African (~750k)
  – Possibly adequate in South Asian, Hispanic

Case/Control WGAA

<table>
<thead>
<tr>
<th></th>
<th>A1</th>
<th>A2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case</td>
<td>( p_{1+} )</td>
<td>( p_{2+} )</td>
</tr>
<tr>
<td>Control</td>
<td>( p_{1-} )</td>
<td>( p_{2-} )</td>
</tr>
</tbody>
</table>
\[ \chi^2 = N(p_{1+}p_{2-} - p_{1-}p_{2+}) \]

Case/Control WGAA

<table>
<thead>
<tr>
<th></th>
<th>A1</th>
<th>A2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case</td>
<td>( p_{11+} )</td>
<td>( p_{12+} )</td>
</tr>
<tr>
<td>Control</td>
<td>( p_{11-} )</td>
<td>( p_{12-} )</td>
</tr>
</tbody>
</table>
\[ \chi^2 = N(p_{11+}p_{22-} - p_{11-}p_{22+}) \]
**Interaction Analysis**

- SNP X SNP
  - Within gene: haplotype
    - Modest interaction space
    - Most haplotype splits do not matter (APOE)
  - Between genes: epistasis
    - Interaction space is vast (500k X 500k)
- SNP X Environment
  - Smaller interaction space (500k X a few environmental measures)

**Limiting the Interaction Space**

- Not all epistatic interactions make sense
  - Physical interactions (lock and key)
  - Physical interactions (subunit stoichiometry)
  - Pathway interactions
  - Regulatory interactions

**Whole Genome Summary**

- Low Hanging Fruit exist (e.g. AMD)
- Tier studies for economic purposes
  - Make sure N is large enough to be powered if all samples were 500k genotyped
- Interactions may be interesting
  - Explore sparingly for hypothesis testing
  - Explore comprehensively for hypothesis generation

**Conclusions**

- Pay attention to study design
  - Sample size
  - Estimated power
  - Multiple Testing
- Analyze SNPs (and haplotypes)
- Keep population structure in mind
- Explore epistasis and environmental interactions after main effects

**Limiting the Interaction Space**

- Not all epistatic interactions make sense
  - Physical interactions (lock and key)
  - Physical interactions (subunit stoichiometry)
  - Pathway interactions
  - Regulatory interactions
Lock and Key

Stoichiometry
E.g. α and β globin in Thalassemia

Pathway
Pathway output can integrate across all steps within the pathway
BUT, many pathways have rate limiting step which can erase upstream variation

Regulatory

Epistasis: SNP X SNP Interactions

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>AC/CC</th>
</tr>
</thead>
<tbody>
<tr>
<td>OR</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>GG</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

Simple model: two dominant loci, two-fold relative risk (RR) to single carriers at either locus, four-fold risk to double carriers. Risk allele frequency 0.05 at both loci.

Tx factor X Tx factor (500 X 500)
Tx factor X gene (10 X 500k)
Epistasis I: Synergistic

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>AC/CC</th>
</tr>
</thead>
<tbody>
<tr>
<td>OR</td>
<td>2.533</td>
<td></td>
</tr>
<tr>
<td>GG</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>GT/TT</td>
<td>2.533</td>
<td>10</td>
</tr>
</tbody>
</table>

Simple model: two dominant loci, two-fold relative risk (RR) to single carriers at either locus, more than four-fold risk to double carriers. Risk allele frequency 0.05 at both loci.

Epistasis II: Permissive

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>AC/CC</th>
</tr>
</thead>
<tbody>
<tr>
<td>OR</td>
<td>1.878</td>
<td></td>
</tr>
<tr>
<td>GG</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>GT/TT</td>
<td>1.878</td>
<td>10</td>
</tr>
</tbody>
</table>

Simple model: two dominant loci, no risk (RR) to single carriers at either locus, more than four-fold risk to double carriers. Risk allele frequency 0.05 at both loci.

Epistasis III: Sufficient

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>AC/CC</th>
</tr>
</thead>
<tbody>
<tr>
<td>OR</td>
<td>1.822</td>
<td></td>
</tr>
<tr>
<td>GG</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>GT/TT</td>
<td>1.822</td>
<td>2</td>
</tr>
</tbody>
</table>

Simple model: two dominant loci, two-fold relative risk (RR) to single carriers at either locus, two-fold risk to double carriers. Risk allele frequency 0.05 at both loci.

Epistasis IV: Exclusive

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>AC/CC</th>
</tr>
</thead>
<tbody>
<tr>
<td>OR</td>
<td>1.733</td>
<td></td>
</tr>
<tr>
<td>GG</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>GT/TT</td>
<td>1.733</td>
<td>2</td>
</tr>
</tbody>
</table>

Simple model: two dominant loci, two-fold relative risk (RR) to single carriers at either locus, no risk to double carriers. Risk allele frequency 0.05 at both loci.

Rare Allele Epistasis

- Main effects are the observed effects analyzing one SNP at a time
- Main effects of rare alleles are not substantially affected by epistatic models
- Are common alleles more substantially affected by epistasis?

Common Allele, No Epistasis

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>AC/CC</th>
</tr>
</thead>
<tbody>
<tr>
<td>OR</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>GG</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>GT/TT</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

Simple model: two dominant loci, two-fold relative risk (RR) to single carriers at either locus, four-fold risk to double carriers. Risk allele frequency 0.3 at both loci (= risk genotype frequency 0.51 at either locus).
Epistasis I: Synergistic

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>AC/CC</th>
</tr>
</thead>
<tbody>
<tr>
<td>OR</td>
<td>4.026</td>
<td></td>
</tr>
<tr>
<td>GG</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>GT/TT</td>
<td>4.026</td>
<td>2</td>
</tr>
</tbody>
</table>

Simple model: two dominant loci, two-fold relative risk (RR) to single carriers at either locus, more than four-fold risk to double carriers. Risk allele frequency 0.3 at both loci.

Epistasis II: Permissive

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>AC/CC</th>
</tr>
</thead>
<tbody>
<tr>
<td>OR</td>
<td>5.59</td>
<td></td>
</tr>
<tr>
<td>GG</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>GT/TT</td>
<td>5.59</td>
<td>1</td>
</tr>
</tbody>
</table>

Simple model: two dominant loci, no risk (RR) to single carriers at either locus, more than four-fold risk to double carriers. Risk allele frequency 0.3 at both loci.

Epistasis III: Sufficient

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>AC/CC</th>
</tr>
</thead>
<tbody>
<tr>
<td>OR</td>
<td>1.325</td>
<td></td>
</tr>
<tr>
<td>GG</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>GT/TT</td>
<td>1.325</td>
<td>2</td>
</tr>
</tbody>
</table>

Simple model: two dominant loci, two-fold relative risk (RR) to single carriers at either locus, two-fold risk to double carriers. Risk allele frequency 0.3 at both loci.

Epistasis IV: Exclusive

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>AC/CC</th>
</tr>
</thead>
<tbody>
<tr>
<td>OR</td>
<td>0.987</td>
<td></td>
</tr>
<tr>
<td>GG</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>GT/TT</td>
<td>0.987</td>
<td>1</td>
</tr>
</tbody>
</table>

Simple model: two dominant loci, two-fold relative risk (RR) to single carriers at either locus, no risk to double carriers. Risk allele frequency 0.3 at both loci.

Main Effects Analysis

- In the vast majority of epistatic models, main effects exist, and point in the right direction
- Epistatic interaction is potentially more important for common alleles
- Limit epistatic exploration to common SNPs with main effects?