Embracing complexity for the genome-wide analysis of imaging phenotypes

Jason H. Moore, Ph.D.
Frank Lane Research Scholar in Computational Genetics
Professor of Genetics and Community and Family Medicine
Dartmouth Medical School, NH
Affiliate Associate Professor of Computer Science
University of New Hampshire, NH
Adjunct Associate Professor of Computer Science
University of Vermont, VT

www.epistasis.org
jason.h.moore@dartmouth.edu

© Jason H. Moore
Introduction

Part 1: Replication of SNPs

Part 2: Replication of Pathways

Part 3: Modeling Complexity
Human Genetic Variation

Equipped with faster, cheaper technologies for sequencing DNA and assessing variation in genomes on scales ranging from one to millions of bases, researchers are finding out how truly different we are from one another.

The unveiling of the human genome almost 7 years ago cast the first faint light on our complete genetic makeup. Since then, each new genome sequenced and each new individual studied has illuminated our genomic landscape in ever more detail. In 2007, researchers came to appreciate the extent to which our genomes differ from person to person and the implications of this variation for deciphering the genetics of complex diseases and personal traits.

Less than a year ago, the big news was triangulating variation between us and our primate cousins to get a better handle on genetic changes along the evolutionary tree that led to humans. Now, we have moved from asking what in our DNA makes us human to striving to know what in my DNA makes me me.
Gene–environment interactions in psychiatry: joining forces with neuroscience

Avshalom Caspi and Terrie E. Moffitt
NIEHS - Personal Toxicology

Environmental/Dietary Exposure → External Contact → Internal Dose → Early Biological Preclinical Response → Phenotype or Clinical

Body of Contact Measurements (Environmental Sensors)

Improved Measures of Body Burden

Biological Measurements (Biological Sensors)

TECHNOLOGY DEVELOPMENT
Inflammation
Circadian Rhythms
Signal Transduction
Neurotransmission

RISK OF SCHIZOPHRENIA

Adapted from Sing et al., (2003)
“One SNP at a time approach”

10% 

Main Effects

90%

Epistasis

Gene-Environment Interaction

Epigenetics

Heterogeneity

“Complex systems approach”
Introduction

Part 1: Replication of SNPs

Part 2: Replication of Pathways

Part 3: Modeling Complexity
False positives in imaging genetics

Andreas Meyer-Lindenberg, a,b,c,d,* Kristin K. Nicodemus, c Michael F. Egan, c,1 Joseph H. Callicott, c Venkata Mattay, b,c and Daniel R. Weinberger c,2

aUnit for Systems Neuroscience in Psychiatry, National Institute for Mental Health, NIH, DHHS, 9000 Rockville Pike, Bethesda, MD 20892-1365, USA
bNeuroimaging Core Facility, National Institute for Mental Health, NIH, DHHS, 9000 Rockville Pike, Bethesda, MD 20892-1365, USA
cClinical Brain Disorders Branch, Genes, Cognition and Psychosis Program, National Institute for Mental Health, NIH, DHHS, 9000 Rockville Pike, Bethesda, MD 20892-1365, USA
dCentral Institute of Mental Health, J5, 68159 Mannheim, Germany
Replicating genotype-phenotype associations

What constitutes replication of a genotype-phenotype association, and how best can it be achieved?

NCI-NHGRI Working Group on Replication in Association Studies

None of these steps should proceed far, however, without conclusive replication of findings from an initial genotype-phenotype association study.
Question

Does failure to replicate a genetic association warrant dismissal?
Hypothesis

Failure to replicate a genetic association may provide important clues about genetic architecture.
Simulation Study
Greene et al., submitted (2009)

- Five different epistasis models
- Six different heritabilities
- N=1600 cases and controls
- Two functional SNPs

- Estimated power to replicate the independent effect of SNP2 after changing allele frequency of interacting SNP1
P(A) = 0.2
P(a) = 0.8

<table>
<thead>
<tr>
<th>SNP 1</th>
<th>AA (0.04)</th>
<th>Aa (0.32)</th>
<th>aa (0.64)</th>
<th>$M_{SNP2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BB (0.25)</td>
<td>0.60</td>
<td>0.47</td>
<td>0.19</td>
<td>0.29</td>
</tr>
<tr>
<td>Bb (0.50)</td>
<td>0.24</td>
<td>0.52</td>
<td>0.44</td>
<td>0.46</td>
</tr>
<tr>
<td>bb (0.25)</td>
<td>0.64</td>
<td>0.21</td>
<td>0.66</td>
<td>0.52</td>
</tr>
<tr>
<td>$M_{SNP1}$</td>
<td>0.43</td>
<td>0.43</td>
<td>0.43</td>
<td></td>
</tr>
</tbody>
</table>
SNPs that fail to replicate should be assessed for gene-gene interaction.
Strength

Novelty

Greene et al., submitted (2009)
Introduction

Part 1: Replication of SNPs

Part 2: Replication of Pathways

Part 3: Modeling Complexity
1) ~500,000 SNPs from the NIMH Study (Sklar et al. 2008)
2) ~500,000 SNPs from the WTCCC (WTCCC 2007)
3) P-value for most significant SNP from each gene recorded
4) Gene ontology (GO) analysis performed using EVA (Reif et al. 2005)
5) Maximum of 500 GO categories considered
6) GO significance assessed using 100,000 permutations
Pathway-Based Analysis: GO Molecular Function

- **NIMH – Detection**
  - Voltage-gated ion channel activity
  - $P = 2.73 \times 10^{-6}$
  - 32/138 genes had SNPs with $P < 0.01$

- **WTCCC – Replication**
  - Voltage-gated ion channel activity
  - $P = 0.01$
  - 22/134 genes had SNPs with $P < 0.01$
Pathway-Based Analysis: GO Molecular Function

- Only 4 genes common to both
  - KCNQ3, KCNMA1, CACNA1E and KCNIP4
  - None were found by NIMH or WTCCC

- **Summary**: Pathways replicate, SNPs don’t

- **Explanation**: Heterogeneity, Epistasis
Introduction

Part 1: Replication of SNPs

Part 2: Replication of Pathways

Part 3: Modeling Complexity
"One SNP at a time approach"

Epistasis
Gene-Environment Interaction
Epigenetics
Heterogeneity

"Complex systems approach"
Y = α + βX
Symbolic Modeling of Epistasis

Jason H. Moore\textsuperscript{a–f} Nate Barney\textsuperscript{a} Chia-Ti Tsai\textsuperscript{g} Fu-Tien Chiang\textsuperscript{g} Jiang Gui\textsuperscript{a, c} Bill C. White\textsuperscript{a}

**GOAL:** make no assumptions about the data and model

Attributes = \{SNP1-SNP500000\}

Functions = \{+ - / * < > = != min max\}

Stochastic search algorithm = genetic programming (GP)
Symbolic Expression Tree

Moore et al., *Human Heredity* (2007)

Accuracy = 0.644
Visualization – Function Mapping

Moore et al., *Human Heredity* (2007)
Interpretation – Information Theory

Moore et al., Human Heredity (2007)
Interpretation – Information Theory

Moore et al., Human Heredity (2007)

Weak Interaction  Strong Interaction

© Jason H. Moore
A Computational Evolution System for Symbolic Modeling

![Diagram of computational evolution system](image)

**Mutation Probability**

- \( P = 0.1 \)

**Mutation Operator**
- \( DeleteOp \)
- \( ChangeOp \)
- \( ChangeOpArg \)
- \( AddOp \)

**Solution Operator**
- ADD
- ADD
- DELETE
- COPY

**Solution**
- >
- !=
- *

**Feedback Loop**

© Jason H. Moore
Artificial Immune Systems
Farmer et al., Physica D (1986)
Greene et al., in preparation (2009)

<table>
<thead>
<tr>
<th>Antibodies</th>
<th>SNP1</th>
<th>SNP2</th>
<th>SNP3</th>
<th>SNP4</th>
<th>SNP5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>= A</td>
<td>!= G</td>
<td>*</td>
<td>= C</td>
<td>!= T</td>
</tr>
<tr>
<td></td>
<td>= T</td>
<td>!= G</td>
<td>= A</td>
<td>*</td>
<td>!= G</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CONTROLS (SELF)</th>
<th>CASES (NON-SELF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A T T T G</td>
<td>C C T T G</td>
</tr>
<tr>
<td>A T T A G</td>
<td>A T T A G</td>
</tr>
<tr>
<td>A T A A C</td>
<td>A C A A G</td>
</tr>
</tbody>
</table>

© Jason H. Moore
Summary and Recommendations

- Psychiatric diseases and imaging phenotypes are complex
- High-dimensional data present many analytical challenges
- Replication is great but lack of replication is meaningless
- Biological pathways must be considered
- Need flexible modeling tools that make no assumptions about the data or models
Acknowledgments

• Graduate Students
  – Richard Cowper, Casey Greene, Krissy Pattin, Sarah Pedergrass, Nadia Penrod, Nima Pouladi, Chantel Sloan, Anna Tyler, Ryan Urbanowicz

• Staff
  – Peter Andrews, Tom Caldwell, Doug Hill, Jeff Kiralis, Pete Schmitt, Paul Thompson

• NIH grants R01 AI59694, R01 LM009012
• jason.h.moore@dartmouth.edu
• compgen.blogspot.com
• www.epistasis.org