Whole genome sequencing and the search for measurable genetic variants influencing behavior and psychiatric illness

Christian Hopfer MD
Associate Professor
UCD School of Medicine
Psychiatry
Overview

- Aggregate Genetic Influences on Psychiatric and Behavioral Traits
- Candidate gene and Linkage studies
- Genome-wide Association studies
- Whole-Genome Sequencing
Mendel 1865 – genetics of discrete traits
R.A. Fisher, 1918

The explanation of quantitative inheritance in Mendelian terms

1 Gene
→ 3 Genotypes
→ 3 Phenotypes

2 Genes
→ 9 Genotypes
→ 5 Phenotypes

3 Genes
→ 27 Genotypes
→ 7 Phenotypes

4 Genes
→ 81 Genotypes
→ 9 Phenotypes

R.A. Fisher, 1918
The explanation of quantitative inheritance in Mendelian terms
Height
Genetics and Behavior

• Evidence for Aggregate Genetic Influences on Behavior (twin/adoption/animal studies)
IQ correlations in Dutch Twins

MZ

DZ
Breeding for Ethanol Sensitivity

Sleep Time Selection

Scores Pooled

Recovery Time, in Seconds

Generation

Long Sleep
Short Sleep
Evidence of Aggregate Genetic Effects on Schizophrenia
Almost every behavioral trait and psychiatric illness measured has been demonstrated to be “heritabile.”

Heritability refers to the percent of the variance with a population that can be explained by genetic influences.

For many behavioral traits and psychiatric disease the heritability is often measured as accounting for 50-80% of the variance.
So…
Let's find the individual, specific, measurable variants or genetic conditions that make up this heritability.

Particularly appealing about this idea, is that we also have a strong theory of causation, i.e., if we find an association between a genetic variant and a trait, we assume that this is a causal association.

Because.. We have rejected Lamarck’s theory of inheritance.
We have some examples of this approach working…

• Many known chromosomal abnormalities associated with specific behavioral syndromes (down’s syndrome, fragile X, 22q11.2 syndrome, etc.)

• The Aldehyde Dehydrogenase point mutation is strongly protective against alcoholism
### TABLE 3: ALDH2*2 Among Alcoholics in Treatment and Controls

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Genotypes of Alcoholics in treatment</th>
<th>Genotypes of Controls</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chen, C '99 Han Chinese</td>
<td>0% had 2 ALDH2*2</td>
<td>4.2% had 2 ALDH2*2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>17% had 1 ALDH2*2</td>
<td>40% had 1 ALDH2*2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>83% had 0 ALDH2*2</td>
<td>55.8% had 0 ALDH2*2</td>
<td>&lt;10^-6</td>
</tr>
<tr>
<td>Higuchi '94 Japanese</td>
<td>8.2% (n=1300) had inactive ALDH2*2</td>
<td>55.8% had 0 ALDH2*2</td>
<td>&lt;10^-6</td>
</tr>
<tr>
<td>Higuchi '96 Japanese</td>
<td>0% had 2 ALDH2*2</td>
<td>6.7% had 2 ALDH2*2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12.2% had 1 ALDH2*2</td>
<td>35.1% had 1 ALDH2*2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>87.8% had no ALDH2*2</td>
<td>58.1% had no ALDH2*2</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
Thomas Hunt Morgan – discoverer of linkage

Linkage analysis
Marker allele $A_1$ cosegregates with dominant disease.
Linkage Markers..
Has this worked for Psychiatric or Behavioral Traits?

• Not well – Linkage findings for psychiatric traits have, in general, not replicated, or not led to clear identification of genetic variants confirmed by subsequent association testing across different samples

• Why?
  Phenotype?
  Effect Size?
Candidate Genes

- Pick a gene
- Have a theory about why it is important
- Test the gene for variants that explain individual differences in your favorite trait, i.e. correlation between variation and trait
- This has been a major focus of research for the past 3 decades
- 1000s of papers have been published with candidate gene findings
• Yet…..
• The field has been characterized by replications and non-replications leading to a lack of clarity about whether variants are actually associated with traits of interest
• The predominant result has been, that after “initial” replications are published, non-replications follow, and, when taking into account unpublished results in meta-analyses, that typically there is not an effect
So far….

• Aggregate genetic effects on behavior and psychiatric illness are demonstrated over and over again through twin and adoption designs as well as animal work
• Cytogenetics demonstrates some clear effects
• There is at least one point mutation that has a fairly large effect (alcohol dehydrogenase)
• Family Pedigrees do not typically demonstrate Mendelian inheritance patterns
• Linkage and Candidate gene approaches do not result in clear findings, despite substantial effort. Individual groups report significant results, but these are not found by other groups, or are inconsistently demonstrated.
The rapidly fading modern era of Genome-Wide Association Studies

Complex disease marker? SNPs are single-base differences in DNA.
How do we test for association?

- We use “tag SNPs” that are correlated (in linkage disequilibrium) with causal variants
High density SNP arrays – up to 5 million SNPs
Genome-Wide Association Studies

500,000 – 5,000,000 SNPs
Human Genome - 3,1 x 10^9 Base Pairs
Genetic Case Control Study

Allele G is ‘associated’ with disease
Published Genome-Wide Associations through 6/2010

904 published GWA at $p \leq 5 \times 10^{-8}$ for 165 traits
GWAS Results of Selected Psychiatric Conditions

• Cigarettes Per Day (Lung Cancer)
• Schizophrenia
• Alcoholism
• Personality
A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25

Rayjean J. Hung1,2*, James D. McKay1*, Valerie Gaborieau1, Paolo Boffetta1, Mia Hashibe1, David Zardze3,

A variant associated with nicotine dependence, lung cancer and peripheral arterial disease

Thorgeir E. Thorgeirsson1*, Frank Geller1*, Patrick Sulem1*, Thorunn Rafnar1*, Anna Wiste1,2,

Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1

Christopher I Amos1, Xifeng Wu1, Peter Broderick2, Ivan P Gorlov1, Jian Gu1, Timothy Eisen3, Qiong Dong1,
Schizophrenia
(Kim et al. 2011, schizophrenia bulletin)

- GWAS studies have implicated variation within the Major Histocompatibility Complex with SCZ
- Additional Loci have been identified that reach genome-wide significance
- A number of these are novel genes or regions that previously had not been known or considered to play a role in the etiology of schizophrenia
- Some Copy Number Variants have been identified as rare but potent risk factors for Schizophrenia
Alcoholism GWAS

Bierut et al. (2010) (n = ~4000)
Treutlein et al. (2009) (n =~ 2000)

No Genome-Wide Significance in Bierut Sample and failure to replicate 2 SNPs that past GWS in Treutlein sample
Personality

- No Genome-wide Significance (Verweijv et al., 2010 Biological Psychiatry)
- Can rule out any individual effect sizes of 1% or greater
GWAS lessons

- Sample sizes need to be at least 10,000 cases and controls, 30,000 seems to be more the “magic” number, and even larger would be better.
- “Extreme” samples may yield additional power.
- Can rule out large and modest effects variants.
- The lack of large or modest effect variants provides an explanation of why linkage and candidate gene studies resulted in inconsistent findings.
- Some additional knowledge has been gained. Much of this knowledge is “negative” knowledge, i.e., we know that candidate genes are not usually showing up in GWAS studies.
- Furthermore…..
Where is the Dark Matter?

Finding the missing heritability of complex diseases

The case of the missing heritability
Hundreds of variants clustered in genomic loci and biological pathways affect human height
Individual genes -HEIGHT

Lango Allen et al. Nature 2010
GIANT Consortium - Height

- 180,000 individuals
- 180 loci identified
- Allelic effect sizes 1 to 4 mm
- Enriched for genes that are connected in biological pathways that underlie skeletal growth
- BUT only ~12% of heritability explained!
Common SNPs explain a large proportion of the heritability for human height

Jian Yang¹, Beben Benyamin¹, Brian P McEvoy¹, Scott Gordon¹, Anjali K Henders¹, Dale R Nyholt¹, Pamela A Madden², Andrew C Heath², Nicholas G Martin¹, Grant W Montgomery¹, Michael E Goddard³ & Peter M Visscher¹

A Commentary on ‘Common SNPs Explain a Large Proportion of the Heritability for Human Height’ by Yang et al. (2010)

Peter M. Visscher¹, Jian Yang¹ and Michael E. Goddard²³
How much of the heritability of psychiatric and behavioral traits is due to small-effect common variants

• Probably a substantial proportion (lubke and boomsma, 2001 submitted)
• But… not all.
• Approximately half of the “heritability” of behavioral traits and psychiatric conditions is accounted for by tagged variants
The Final Frontier

Whole Genome Sequencing
~$1 Billion in 2000

Venter
Clinton
Collins
The complete genome of an individual by massively parallel DNA sequencing

David A. Wheeler1*, Maithreyan Srinivasan2*, Michael Egholm2*, Yufeng Shen1*, Lei Chen1, Amy McGuire3, Wen He2, Yi-Ju Chen2, Vinod Makhijani2, G. Thomas Roth2, Xavier Gomes2, Karrie Tartaro2†, Faheem Niazi2, Cynthia L. Turcotte2, Gerard P. Irzyk2, James R. Lupski4,5,6, Craig Chinault4, Xing-zhi Song1, Yue Liu1, Ye Yuan1, Lynne Nazareth1, Xiang Qin1, Donna M. Muzny1, Marcel Margulies2, George M. Weinstock1,4, Richard A. Gibbs1,4 & Jonathan M. Rothberg2†

The association of genetic variation with disease and drug response, and improvements in nucleic acid technologies, have given great optimism for the impact of ‘genomic medicine’. However, the formidable size of the diploid human genome, approximately 6 gigabases, has prevented the routine application of sequencing methods to deciphering complete individual human genomes. To realize the full potential of genomics for human health, this subject’s DNA, including single nucleotide polymorphisms (SNPs), small insertions and deletions (indels), and copy number variation (CNV).

The 454 base-calling software provides error estimates (Q values) for each base. We developed a three-step filtering process using the patterns of error and associated Q values from the 454 base-calling software to improve the accuracy of SNP discovery. An initial 14 mil
~$5000 in 2011

Systems / HiSeq 2000

HiSeq 2000

The HiSeq 2000 sequencing system offers unprecedented output and a breakthrough user experience. Leveraging Illumina's proven and widely-adopted, reversible terminator-based sequencing by synthesis chemistry in combination with innovative engineering, HiSeq 2000 delivers the industry's highest sequencing output and fastest data generation rate. Human interaction design features and the easiest sequencing workflow set a new standard for simplicity and user experience.

Applications: DNA Sequencing, Gene Regulation Analysis, Sequencing-Based Transcriptome Analysis, SNP Discovery and Structural Variation Analysis, Cytogenetic Analysis
Whole – Genome Sequencing is feasible NOW

What are the technical aspects involved in conducting behavioral and psychiatric research utilizing this technology?

What can we expect?

Study design?