Whole Genome Analyses (WGA)

John Hardy (hardyj@mail.nih.gov)
Andy Singleton (singleta@mail.nih.gov)

Laboratory of Neurogenetics, NIA/NIH, Building 35, Bethesda Campus
Whole Genome Study For Alzheimer’s Disease

500K Affymetrix Array

(presently underpowered ~180 cases and controls: collaboration with TGen: now 800 cases and controls: data analysis ongoing)
Our Studies (Illumina Bead Station)

- **Whole Genome Association Analysis of Parkinson’s Disease**
  - 276 Cases and Controls NIA/NINDS funded (completed: in follow up)
- **Whole Genome Association Analysis of Ischemic Stroke**
  - 276 Cases and Controls NIA/NINDS funded (completed: in follow up)
- **Whole Genome Association Analysis of ALS**
  - 276 Cases and Controls NIA/NINDS/ALSA funded (in progress)
- **Whole Genome Analysis of Haplotypic Brain Expression**
  - 300 Control Brains NIA/TGen funded (in progress)
- **Whole Genome Analysis of African Americans**
  - 200 from the HANDLs Study (Baltimore, NIA/Michele Evans PI) (in progress: more planned)
Three Surprises:-

• Data quality: routinely >99% of data
• In North American Caucasian Controls, ~10% showed extensive homozygosity (not true of African Americans)
  – (parents were ~2\textsuperscript{nd}-4\textsuperscript{th} cousins)
• In North American Controls, ~9% had significant structural variability (large insertions and deletions)
  – (some was cell line specific, but much was not: what is “normal”?)

• Realized both homozygosity and structural variability could be disease-related

http://ccr.coriell.org/ninds/
Loss of Heterozygosity

(10% North American Controls Show Evidence for Consanguinity)
Detection of homozygous parkin deletion causing Parkinson’s disease
Homozygosity Mapping on Infinium 300K: one hit linkage for a new disease

- Recessive young onset ataxia
- Single segregating region ~1Mb
Structural Alterations
Chromosome 5 Control Male of 65 years
APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy

Anne Rovelet-Lecruix¹, Didier Hannequin¹,², Gregory Raux¹, Nathalie Le Meur³, Annie Laquerrière⁴, Anne Vital⁵, Cécile Dumanchin¹, Sébastien Feuillette¹, Alexis Brice⁶, Martine Vercelletto⁷, Frédéric Dubas⁸, Thierry Frebourg¹ & Dominique Campion¹,⁹
BREVIA

α-Synuclein Locus Triplication
Causes Parkinson’s Disease

A. B. Singleton,† M. Farrer,‡ J. Johnson,§ A. Singleton,¶ S. Hague,¶ J. Kachergus,¶ M. Hulihan,¶ T. Peuralinna,∥ A. Dutra,∥ R. Nussbaum,∥ S. Lincoln,∥ A. Crawley,∥ M. Hanson,∥ D. Maraganore,∥ C. Adler,∥ M. R. Cookson,§ M. Muenzer,∥ M. Baptista,∥ D. Miller,∥ J. Blattner,∥ J. Hardy,∥ K. Gwinn-Hardy∥
Whole Genome Association Analyses

• Population choice is important
• Data handling is not trivial
  – (our lab has been generating ~6,000,000 a day for 6 months) and now has
    ~1,000,000,000 genotypes)
Population Choice: LRRK2 and PD

<table>
<thead>
<tr>
<th>Population Choice: LRRK2 and PD</th>
</tr>
</thead>
</table>

Genetic screening for a single common LRRK2 mutation in familial Parkinson’s disease

William Cheah, Pihan Poon, David Kam, Yoko Tase, Carin Ellegaard, Helen McWilliams, Tony Reed, Alice Bulloch, Clifford Fujita, Andrew Singleton, Ted Davis, for the Parkinson Study Group, PDGene: Parkinson disease genetic initiative.

A common LRRK2 mutation in idiopathic Parkinson’s disease


A frequent LRRK2 gene mutation associated with autosomal dominant Parkinson’s disease


- Gly2019Ser, alters conserved aa within the kinase activation loop
- Could this be constitutively active?
- Involved in target recognition?

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>UK BRAIN BANK</td>
<td>1.6%</td>
<td>5%</td>
<td>2.8%</td>
<td>1%</td>
<td>11%</td>
<td>23%</td>
<td>39%</td>
</tr>
<tr>
<td>NORTH AMERICAN SIB-PAIR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NORTH AMERICAN FAMILIAL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NORTH AMERICAN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PORTUGUESE CLINIC-BASED</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASHKENAZI JEWISH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NORTH AFRICAN ARAB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Report
Identification of a Novel LRRK2 Mutation Linked to Autosomal Dominant Parkinsonism: Evidence of a Common Founder across European Populations


LRRK2 mutations and Parkinsonism

Mathias Solt, Ignacio F. Mata, Jennifer MKucharczyk, Owen A. Ross.
Whole Genome Analyses of 276 US PD cases

Raphael Gibbs; gibbsr@mail.nih.gov
LNG Data Management Software

**GERON** (Raph Gibbs, Andy Singleton):
- Clinical data, sample tracking and genotype storage

**SNP GWA** (Carl Langefeld, Matt Stiegert at Wake Forest):
- Analysis (Hardy Weinberg; dominant; recessive; additive; haplotypic)
Whole Genome Data

• Association… sure…
  – Probably OK down to ~ $\lambda$s of 1.5-2.0
• Homozygosity mapping
  – In kindreds and in populations
• Insertion/deletion cataloguing
  – What is “normal” and what is pathologic?
• Genetic Ancestry
  – Different populations
• Whole genome diplotype/expression correlation
  – 300 Human Cortical Samples (500K Affy SNP data from TGen: 24K Illumina Expression data)
• Cell lines
  – What are (stem) cell lines like? (a mess)
Whole Genome Analyses

- 1) Whole genome associations will pick up alleles of large effect
- 2) Tell you what is NOT there (valuable to NIH) (within the bounds of the study design and population of course)
- 3) They are ADDITIVE: allowing studies to be pooled easily
- 4) Identify quickly and easily insertions/deletions (need a normal catalogue)
- 5) Enables homozygosity mapping in “outbred” populations (parkin example)
- 6) Enables cell lines to be characterized (stem cells etc)
- 7) Enables genotype/expression correlations for cis/(trans?) correlations of gene expression (valuable for complex trait genetic associations)
Many Genetic Associations Likely to Reflect Differences in Expression:-

• Three Common Haplotypes of Gene $\omega$
  $\omega$A gives ↑
  $\omega$B gives →
  $\omega$C gives ↓

• 300 Human Control Cortices: Fully Genotyped (Affy 500K array): and Full Expression Array (Illumina 24K expression Array)

• If disease is associated with High Expression, genetic association should be seen in $\omega$AA homozygotes and protection in $\omega$CC with $\omega$BB intermediate
Need to have proper diversity to know what is normal and to find admixture (CEPH Diversity series)

Distribution of Tau Haplotype
Common cell lines

M17

HEK293
Federally approved Stem Cell Lines
The Future

- All genetic samples will have WG data (now ~$700 a sample)
- All cell lines will have WG data
- Need a catalogue of variability
- Need real diversity
- Perhaps start with homogenous cohorts for whole genome association analyses