Update on Genome-Wide Association Studies:
We Live in Interesting Times

U.S. Department of Health and Human Services
National Institutes of Health
National Human Genome Research Institute

Teri A. Manolio, M.D., Ph.D.
Director, Office of Population Genomics
Senior Advisor to the Director, NHGRI,
for Population Genomics
September 19, 2007
We Live in Interesting Times…

“‘May he live in interesting times.’ Like it or not we live in interesting times.”

--Robert Kennedy, June 7, 1966

May you come to the attention of those in authority.

May you find what you are looking for.

Wikipedia, accessed 11Sep07
May You Live in Interesting Times…

Since 2005, over 30 genome-wide association studies have identified robust associations with genetic variants for nearly 20 common, complex diseases and traits:

- 10 Mar 2005: Age-related macular degeneration
- 30 Apr 2006: QT interval prolongation
- 19 Oct 2006: Neovascular AMD
- 26 Oct 2006: Inflammatory bowel disease
- 11 Feb 2007: Type 2 diabetes
- 5 Mar 2007: Crohn’s disease
- 12 Apr 2007: Obesity
Genome-wide association study of prostate cancer identifies a second risk locus at 8q24

Meredith Yeager1,2, Nick Orr3, Richard B Hayes2, Kevin B Jacobs4, Peter Kraft5, Sholom Wacholder2, Mark J Minichiello6, Paul Fearnhead7, Kai Yu2, Nilanjan Chatterjee2, Zhaoming Wang1,2, Robert Welch1,2, Brian J Staats1,2, Eugenia E Calle8, Heather Spencer Feigelson8, Michael J Thun8, Carmen Rodriguez8, Demetris Patsopoulos9, Walter Willet9, Edward Giovannucci9, David J Clark1, Christopher A Haiman1, Nick Patterson2, Matthew L Freedman2,3, Simon R Myers2, Malcolm C Pike1, Alicja Waliszewska2,4,5, Julie Neubauer2,4, Arti Tandon2,4, Christine Schirmer2,4, Gavin J McDonald2,4, Steven D Long2,4, David Witting2,4, Kathleen Cho Ch программа 3,4,5, Brian F O’Roak2,4,5, Julius Gudmundsson1,17, Patrick Sulem1,17, Andrei Manolescu1,17, Laufey T Amundadottir1,17, Daniel Gudbjartsson1, Agnar Helgason1, Thorunn Rafnar1, Jon T Berghthorsson1, Bjarni A Aagnarsson2, Adam Baker1, Asgeir Sigurdsson1, Kristrun R Benediktsdottir2, Margret Jakobsdottir1, Jianfeng Xu3, Thorarinn Blondal1, Jelena Kostic1, Jielin Sun3, Shyamali Ghosh1, Simon N Stacey1, Magali Mouy1, Jona Saemundsdottir1, Valgerdur M Backman1, Kristleifur Kristjansson1, Alejandro Tres4,7, Alan W Partin5, Marjo T Albers-Ackers5, Javier Godino-Ivan Marcos7, Patrick C Walsh5, Dorine W Swinkels8, Sebastian Navarrete9, Sarah D Isaacs5, Katja K Aben10, Theresa Graif11, John Cashy11, Manuel Ruiz-Echarri4,
A Genome-Wide Association Study of Type 2 Diabetes in Finns Detects Multiple Susceptibility Variants

Genome-Wide Association Analysis Identifies Loci for Type 2 Diabetes and Triglyceride Levels

Replication of Genome-Wide Association Signals in UK Samples Reveals Risk Loci for Type 2 Diabetes

Eleftheria Zeggini, Michael N. Weedon, Cecilia M. Lindgren, Timothy M. Frayling, Katherine S. Elliott, Hana Lango, Nicholas J. Timpson, John R. B. Perry
A Common Allele on Chromosome 9 Associated with Coronary Heart Disease

Ruth McPherson,1*† Alexander Pertsemlidis,2* Nihan Kavaslar,1 Alexandre Stewart,1 Robert Roberts,1 David R. Cox,3 David A. Hinds,3 Len A. Pennacchio,4,5 Anne Tybjaerg-Hansen,6 Aaron R. Folsom,7 Eric Boerwinkle,8 Helen H. Hobbs,2,9 Jonathan C. Cohen2,10†

A Common Variant on Chromosome 9p21 Affects the Risk of Myocardial Infarction

Anna Helgadottir,1* Gudmar Thorleifsson,1* Andrei Manolescu,1* Solveig Gretarsdottir,1 Thorarinn Blondal,1 Aslaug Jonasdottir,1 Adalbjorg Jonasdottir,1 Asgeir Sigurdsson,1 Adam Baker,1 Arnar Palsson,1 Gisli Masson,1 Daniel F. Gudbjartsson,1 Kristinn P. Magnussson,1 Karl Andersen,2 Allan I. Levey,3 Valgerdur M. Backman,1 Sigurborg Matthiasdottir,1 Thorbjorg Jonsdottir,1 Stefan Palsson,1 Helga Einarsdottir,1 Steinunn Gunnarsdottir,1
Genome-wide association study identifies novel breast cancer susceptibility loci

A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer

Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor–positive breast cancer

Simon N Stacey¹, Andrei Manolescu¹, Patrick Sulem¹, Thorunn Rafnar¹, Julius Gudmundsson¹, Sigurjon A Gudjonsson¹, Gisli Masson¹, Margret Jakobsdottir¹, Steinunn Thorlacius¹, Agnar Helgason¹, Katja K Aben²,³, Luc J Strobbe⁴, Marjo T Albers-Akkers⁵, Dorine W Swinkels⁶, Brian E Henderson⁶, Laurence N Kolonel⁷, Loic Le Marchand⁷, Esther Millastre⁸, Raquel Andres⁸, Javier Godino⁹, Maria Dolores Garcia-Prats¹⁰, Eduardo Polo¹¹, Alejandro Tres⁸, Magali Mouy¹, Jona Saemundsdottir¹, Valgerdur M Backman¹, Larus Gudmundsson¹, Kristleifur Kristjansson¹, Jon T Bergthorsson¹, Jelena Kostic¹, Michael L Frigge¹, Frank Geller¹, Daniel Gudbjartsson¹, Helgi Sigurdsson¹², Thora Jonsdottir¹²,
Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls

We followed up on 37 SNPs from 31 distinct loci, associated at $P < 10^{-5}$ on initial analysis of the WTCCC data set. Support for some of these markers diminished in the final WTCCC analysis after extensive data filtering. We selected two markers for each locus where low linkage disequilibrium (LD) between associated SNPs in

Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes


Nature and Nature Genetics, 7Jun2007
2007: The Year of GWA Studies?

Consistently replicated associations found for:

- 10 Jun 2007: Celiac disease
- 1 Jul 2007: Atrial fibrillation
- 8 Jul 2007: Colorectal cancer
- 15 Jul 2007: Gallstones
- 18 Jul 2007: Periodic limb movements in sleep
- 19 Jul 2007: HIV viral setpoint
- 26 Jul 2007: Childhood asthma
- 29 Jul 2007: Multiple sclerosis
- 1 Aug 2007: Amyotrophic Lateral Sclerosis
- 9 Aug 2007: Exfoliation glaucoma
- 2 Sep 2007: Height
- 5 Sep 2007: Rheumatoid arthritis
- 18 Sep 2007: ??
What is a GWA Study?

• Method for interrogating all 10 million variable points across human genome
• Variation inherited in groups, or blocks, so not all 10 million points have to be tested
• Blocks are shorter (so need to test more points) the less closely people are related
• Technology now allows studies in unrelated persons, assuming ~10,000 base pair lengths in common (300,000 - 500,000 markers)
SNPs 1 / 300 bases
Mapping the Relationships Among SNPs

# Distances Among East Coast Cities

<table>
<thead>
<tr>
<th></th>
<th>Providence</th>
<th>New York</th>
<th>Philadelphia</th>
<th>Baltimore</th>
<th>Washington</th>
</tr>
</thead>
<tbody>
<tr>
<td>Providence</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New York</td>
<td>210</td>
<td>152</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Philadelphia</td>
<td>320</td>
<td>237</td>
<td>86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baltimore</td>
<td>430</td>
<td>325</td>
<td>173</td>
<td>87</td>
<td></td>
</tr>
<tr>
<td>Washington</td>
<td>450</td>
<td>358</td>
<td>206</td>
<td>120</td>
<td>34</td>
</tr>
</tbody>
</table>

- **< 100**
- **101-200**
- **201-300**
- **301-400**
- **> 400**
Distances Among East Coast Cities

- Providence: 59
- New York: 210
- Philadelphia: 320
- Baltimore: 325
- Washington: 358

- Providence
  - New York: 152
  - Philadelphia: 237
- New York
  - Providence: 210
  - Philadelphia: 173
- Philadelphia
  - Providence: 320
  - New York: 86
- Baltimore
  - Providence: 325
  - New York: 173
  - Philadelphia: 87
- Washington
  - Providence: 358
  - New York: 206
  - Philadelphia: 120
  - Baltimore: 34

Distance ranges:
- < 100
- 101-200
- 201-300
- 301-400
- > 400
Distances Among East Coast Cities

- Boston
- Providence
- New York
- Philadelphia
- Baltimore
- Washington
Distances Among East Coast Cities

- Boston
- Providence
- New York
- Philadelphia
- Baltimore
- Washington
Mapping the Relationships Among SNPs

One SNP May Serve as Proxy for Many

Progress in Genotyping Technology

- ABI TaqMan
- ABI SN Plex
- Illumina Golden Gate
- Illumina
- Affymetrix 10K
- Affymetrix
- Affymetrix 100K/500K
- Perlegen
- Infinium/Sentrix
- MegAllele

Number of SNPs:
- 2001
- 2005

Cost per genotype (Cents, USD):
- 10
- 10^2

Courtesy S. Chanock, NCI
Continued Progress in Genotyping Technology

Cost per person (USD)


Affymetrix 500K

Illumina 317K

Illumina 550K

Illumina 650Y

Courtesy S. Gabriel, Broad/MIT
## Cost of a Genome-Wide Association Study in 2,000 People

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of SNPs</th>
<th>Cost/SNP</th>
<th>Cost/Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>10,000,000</td>
<td>$1.00</td>
<td>$20 billion</td>
</tr>
<tr>
<td>2007</td>
<td>500,000</td>
<td>0.1¢</td>
<td>$1 million</td>
</tr>
</tbody>
</table>
### GWA Genotyping Data, Chromosome 22, Parkinson’s Study

<table>
<thead>
<tr>
<th>Study ID</th>
<th>Case/Control Status</th>
<th>rs5747620</th>
<th>rs2236639</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Allele 1</td>
<td>Allele 2</td>
</tr>
<tr>
<td>14</td>
<td>Case</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>20</td>
<td>Case</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>41</td>
<td>Case</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>412</td>
<td>Control</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>592</td>
<td>Control</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>665</td>
<td>Control</td>
<td>T</td>
<td>C</td>
</tr>
</tbody>
</table>

# Association of rs2236639 Alleles with Development of Parkinson Disease (Made Up!)

<table>
<thead>
<tr>
<th>Variant Allele (A)</th>
<th>Develop Disease</th>
<th>Do Not Develop Disease</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present</td>
<td>10</td>
<td>70</td>
<td>80</td>
</tr>
<tr>
<td>Absent</td>
<td>40</td>
<td>880</td>
<td>920</td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
<td>950</td>
<td>1,000</td>
</tr>
</tbody>
</table>

Relative Risk = \( \frac{\text{Risk in Exposed}}{\text{Risk in Unexposed}} \) = \( \frac{10/80}{40/920} \) = \( \frac{12.5\%}{4.3\%} \) = 2.9
P Values of GWA Scan for Age-Related Macular Degeneration

Genome-Wide Scan for Type 2 Diabetes in a Scandinavian Cohort

http://www.broad.mit.edu/diabetes/scandinavs/type2.html
Genome-Wide Scan for Crohn Disease in Belgian Cases and Controls

Genome-Wide Scan for Type 2 Diabetes in French Case-Control Study

Wellcome Trust Genome-Wide Association Study of Seven Common Diseases

Genome-Wide Scan for Breast Cancer in Postmenopausal Women

Genome-Wide Scan for Coronary Heart Disease in Ottawa Case-Control Study

Genome-Wide Scan for Sporadic Amyotrophic Lateral Sclerosis

Genome-Wide Scan for Prostate Cancer

Association Analysis of SNPs across FGFR2

Lessons Learned from Initial GWA Studies

<table>
<thead>
<tr>
<th>Signals in Gene “Deserts”</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Prostate Cancer</td>
<td>8q24</td>
</tr>
<tr>
<td>Crohn’s Disease</td>
<td>5p13.1, 1q31.2, 10p21</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Signals in Common</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Diabetes, CHD, Melanoma</td>
<td>CDKN2A/2B</td>
</tr>
<tr>
<td>Prostate, Breast, Colon Cancer</td>
<td>8q24 region</td>
</tr>
<tr>
<td>Crohn’s Disease, Psoriasis</td>
<td>IL23R</td>
</tr>
<tr>
<td>Crohn’s Disease, T1DM</td>
<td>PTPN2</td>
</tr>
</tbody>
</table>
“There have been few, if any, similar bursts of discovery in the history of medical research...”
Unique Aspects of GWA Studies

• Permits examination of inherited genetic variability at unprecedented level of resolution
• Permits "agnostic" genomewide comparison
• Most robust associations in GWA studies have not been with genes previously suspected of being related to the disease
• Some associations in regions not even known to harbor genes

“The chief strength of the new approach also contains its chief problem: with more than 500,000 comparisons per study, the potential for false positive results is unprecedented.”

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

The study of human genetics has recently undergone a dramatic transition with the completion of both the sequencing of the human genome and the mapping of human haplotypes of the most common form of genetic variation, the single nucleotide polymorphism (SNP)\textsuperscript{1-3}. In concert with this rapid expansion of detailed genomic information, cost-effective genotyping technologies have been developed that can assay hundreds of thousands of SNPs simultaneously. Together, these advances have allowed a systematic, even ‘agnostic’, approach to genome-wide interrogation, thereby relaxing the requirement for strong prior hypotheses.

So far, comprehensive reviews of the published literature, most of which reports work based on the candidate-gene approach, have demonstrated a plethora of questionable genotype-phenotype associations, replication of which has often failed in independent studies\textsuperscript{4-7}. As the transition to genome-wide association studies occurs, the challenge will be to separate true associations from the blizzard of false positives attained through attempts to rep-studies because of issues in either the initial study or the attempted replication\textsuperscript{8-10}. Small sample size is a frequent problem and can result in conclusion from the literature because follow-up studies have not consistently analysed the same markers or those in perfect linkage dis-

Flow of Investigation: From Genome-Wide Association to Clinical Translation

COMPONENT | PERCENT
--- | ---
Initial Genome-Wide Association (GWA) Studies | 30-40
Replication/Fine Mapping | 12-15
Sequencing/Genotyping | 10-12
Functional Studies | 5-10
Translational Studies | 5-10
Data Analysis | 12-15
Database | 5-10
Availability of GWA Data in NIH Databases: Current

- Cancer Biomedical Information Grid (caBIG) and Cancer Genetic Markers of Susceptibility (CGEMS): https://caintegrator.nci.nih.gov/cgems/
Possible Implications of Many Variants of Small Effect

- Need not carry all of them to develop disease
- Probably need to carry more than one, unless very strong environmental interaction
- Some may affect same pathways and be duplicative
- Others may affect different pathways, so some key combination(s) needed
- Should be possible to identify “clusters” of variants carried by different groups of cases
- May be possible to classify on molecular basis
“The more we find, the more we see, the more we come to learn.
The more that we explore, the more we shall return.”

Sir Tim Rice, *Aida*, 2000