A Genome-Wide Association Study of Type 2 Diabetes

Michael Boehnke
Department of Biostatistics
Center for Statistical Genetics
University of Michigan

for the FUSION, CIDR, DGI, WTCCC/UKT2D, and SardiNIA Study Investigators

GAIN Meeting, Bethesda, Maryland
October 18, 2007
Introduction

• Genome-wide association (GWA) studies seek to identify genetic variants that predispose to human diseases, influence (disease-related) quantitative traits

• GWAs enabled by catalogs of genetic variation, SNP genotype arrays, drop in genotype costs

• Why GWAs?
  – better understand disease etiology
  – identify targets for drug development, tailoring of drug therapies
  – predict disease risk
  – for complex traits, more effective than linkage, candidate gene studies

• GWAs have now identified many disease-predisposing variants
Progress in the identification of gene variants for common diseases

<table>
<thead>
<tr>
<th>Disease</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Year 5</th>
<th>Year 6</th>
<th>Year 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cholesterol</td>
<td>2000</td>
<td>2001</td>
<td>2002</td>
<td>2003</td>
<td>2004</td>
<td>2005</td>
<td>2006</td>
</tr>
<tr>
<td>Obesity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2007</td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2005</td>
</tr>
<tr>
<td>QT interval</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2006</td>
</tr>
<tr>
<td>Atrial Fibrillation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2007</td>
</tr>
<tr>
<td>Type 2 Diabetes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2003</td>
</tr>
<tr>
<td>Prostate cancer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2007</td>
</tr>
<tr>
<td>Breast cancer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2005</td>
</tr>
<tr>
<td>Colon cancer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2006</td>
</tr>
<tr>
<td>Age Related Macular Degeneration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2006</td>
</tr>
<tr>
<td>Crohns Disease</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2006</td>
</tr>
<tr>
<td>Type 1 Diabetes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2003</td>
</tr>
<tr>
<td>Systemic Lupus Erythematosus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2006</td>
</tr>
<tr>
<td>Asthma</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2006</td>
</tr>
<tr>
<td>Restless leg syndrome</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2003</td>
</tr>
<tr>
<td>Gallstone disease</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2006</td>
</tr>
<tr>
<td>Multiple sclerosis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2003</td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2006</td>
</tr>
<tr>
<td>Glaucoma</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2004</td>
</tr>
</tbody>
</table>

Slide from David Altshuler
Outline of Presentation

• FUSION study of T2D

• Design, QC, initial results of FUSION/CIDR T2D GWA

• Results of initial meta-analysis of FUSION, DGI, WTCCC/UKT2D GWA studies

• Current follow-up for T2D with DGI, WTCCCI

• GWAs and follow-up for T2D-related traits (SardiNIA, DGI, others)
FUSION Study: Finland-United States Investigation of NIDDM Genetics

National Public Health Institute, Helsinki (Jaakko Tuomilehto)
USC Keck School of Medicine, Los Angeles (Richard Bergman)
National Human Genome Research Institute, Bethesda (Francis Collins)
University of Michigan School of Public Health (Michael Boehnke)
University of North Carolina School of Medicine (Karen Mohlke)
FUSION Study Goals

Identify genetic variants that predispose to type 2 diabetes (T2D) or are responsible for variability in T2D-related quantitative traits (QTs)
Why (or why not) GWA of T2D?

- T2D huge, growing public health problem worldwide
- T2D strongly familial
  - T2D MZ twin concordance rate ~2x DZ rate
  - T2D risk to 1° relatives 3-4x population risk
- Despite much effort, as of March 2007 clear consensus on only three T2D loci: *PPARG*, *KCNJ11*, *TCF7L2*; and associated risks modest
- J. V. Neel: “the geneticist’s nightmare”
FUSION Study Design

- Families ascertained through T2D affected sibling pairs (ASPs)
- All available affected sibs, parents
- Some spouses, offspring

- More recently, unrelated T2D cases, NGT controls from
  - Finrisk 1987, 2002; D2D; Health 2000; Action LADA
  - Savitaipale
Current FUSION Study Samples

Affected sib pair (ASP) families:
F1: 1129 T2D cases in 580 families
F2: 580 T2D cases in 275 families

Stage 1 association samples:
Familial and pop-based cases 1161
Spouses and pop-based controls 1174

Stage 2 association samples:
Population-based cases 1215
Population-based controls 1258
FUSION Genomewide Association Study

• Stage 1: Genotyped on Illumina 317K chip (CIDR)

• Stage 2: Genotyping on best GWA SNPs (Bethesda, Chapel Hill)
  – SNPs associated with T2D or related traits
  – consider also genome annotation
  – >100 now, GWA soon (CIDR)

• 80% power to detect at genome-wide significance:
  – Stage 1: \( OR = 1.4-1.5 \) (depending on MAF)
  – Stage 1 + 2: \( OR = 1.3-1.4 \)
FUSION Stage 1 Genotyping and QC (1)

• 317,503 SNPs genotyped on Illumina HumanHap300 BeadChip; CIDR pilot project

• Included 121 trios, 79 duplicate samples

• QC based on HWE, data completeness, duplicate and Mendel errors

• SNP exclusion, flagging, review
FUSION Stage 1 Genotyping and QC (2)

• 1,808 SNPs (0.6%) excluded from analysis
  Hardy-Weinberg equilibrium p-value < $10^{-6}$
  < 90% successful genotypes
  > 3 Mendelian or duplicate errors
  < 10 minor alleles

• 4,881 SNPs (1.5%) flagged for analysis
  Hardy-Weinberg equilibrium p-value < $10^{-4}$
  < 95% successful genotypes
  >1 Mendelian or duplicate error
Genotyping Quality for 315,635 SNPs

- Successfully genotyped samples: 99.7% (with call frequency > 97.5%)
- Successfully called genotypes: 99.84%
- Duplicate consistency rate (79 pairs): 99.996%
- Mendelian consistency rate (121 trios): 99.97%
FUSION T2D GWA Results

1161 Finnish T2D cases + 1174 Finnish NGT controls

Logistic regression: additive model adjusted for age, gender, birth province
FUSION GWA Results: Known Positives

1161 Finnish T2D cases + 1174 Finnish NGT controls
### Excess of Strongly Associated SNPs?

<table>
<thead>
<tr>
<th>Threshold</th>
<th>Expected Number</th>
<th>Observed Number</th>
<th>Empirical p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>p-value &lt; $10^{-6}$</td>
<td>0.3</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>p-value &lt; $10^{-5}$</td>
<td>3</td>
<td>3</td>
<td>0.54</td>
</tr>
<tr>
<td>p-value &lt; $10^{-4}$</td>
<td>31</td>
<td>43</td>
<td>0.19</td>
</tr>
</tbody>
</table>

Empirical p-value obtained by 100 permutations
Population Stratification?

• Differences in SNP allele frequencies across Finland (e.g. Willer et al. 2005)

• Cases, controls frequency matched: birth province, age, sex

• Logistic regression analysis genomic control $\lambda_{GC} = 1.026$

• QQ plot of p-values looks like a straight line

• Conditional logistic regression constructing matched sets of cases, controls based on IBS sharing gave similar results
Next Steps

• Imputation of non-genotyped HapMap SNPs
• Genotyping of Stage 2 samples
• Meta-analysis and follow up with DGI, UKT2D/WTCCC
• Genome-wide analysis of T2D-related traits
• Fine mapping, resequencing, functional genomics
Imputation of Non-Genotyped SNPs (1)

• Used our genotypes, HapMap CEU genotypes to impute genotypes for all HapMap common SNPs in FUSION using MACH (Li et al. 2007)

• Goal 1: test for association with more of common SNPs in genome ("better coverage")

• Goal 2: allow easier combination of results across genotyping platforms (e.g. Illumina 317K, Affy 500K)
Imputation of Non-Genotyped SNPs (2)

- Imputed ~2.15 million HapMap SNPs with minor allele frequency (MAF) >1% in FUSION
- 2.09 million of these SNPs passed QC
- Increased coverage at $r^2 > .8$ of HapMap SNPs with MAF >1% from 78% to 89%
## Imputed vs. Genotyped SNP Results

<table>
<thead>
<tr>
<th>Allele frequency</th>
<th>P-value</th>
<th>Odds ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Imputed</td>
<td>Genotyped</td>
</tr>
<tr>
<td>Imputed</td>
<td></td>
<td>Genotyped</td>
</tr>
<tr>
<td>.024</td>
<td>.021</td>
<td>2.5 x 10^{-6}</td>
</tr>
<tr>
<td>.543</td>
<td>.540</td>
<td>5.3 x 10^{-6}</td>
</tr>
<tr>
<td>.114</td>
<td>.136</td>
<td>2.0 x 10^{-5}</td>
</tr>
<tr>
<td>.494</td>
<td>.490</td>
<td>6.6 x 10^{-5}</td>
</tr>
<tr>
<td>.927</td>
<td>.924</td>
<td>7.5 x 10^{-5}</td>
</tr>
<tr>
<td>.744</td>
<td>.753</td>
<td>1.4 x 10^{-4}</td>
</tr>
<tr>
<td>.289</td>
<td>.291</td>
<td>1.7 x 10^{-4}</td>
</tr>
<tr>
<td>.970</td>
<td>.973</td>
<td>1.9 x 10^{-4}</td>
</tr>
<tr>
<td>.401</td>
<td>.361</td>
<td>6.3 x 10^{-4}</td>
</tr>
<tr>
<td>.817</td>
<td>.816</td>
<td>9.5 x 10^{-4}</td>
</tr>
<tr>
<td>.605</td>
<td>.605</td>
<td>9.9 x 10^{-4}</td>
</tr>
</tbody>
</table>
Genotyping of Stage 2 Samples

- Test SNPs strongly associated in FUSION Stage 1

- Advantage SNPs based on annotation:
  - non-synonymous SNPs
  - critical splice variants
  - candidate genes, conserved regions, linkage

- ~30 SNPs followed up from FUSION Stage 1 alone
Results of Initial Stage 2 Genotyping

-log_{10}(p-value)

SNP genotyped in stage 2

Stage 1
Stage 2
Stage 1+2

TCF7L2
PPARG
KCNJ11
Chr 11
SLC30A8
IGF2BP2
Chr 7

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28
For “geneticist’s nightmare”, more samples needed

Diabetes Genetic Initiative (DGI): Finnish, Swedish T2D cases, non-DM controls; some from discordant sibships

WTCCC/UKT2D: unrelated UK T2D cases, random controls

Genotyped Affymetrix 500K; ~380K usable SNPs

Meta-analysis combined ORs using precision-weighted combination of results → follow up
Three Collaborating Studies

# cases + # controls

- **FUSION**
  - 1161 + 1174
  - 1215 + 1258

- **DGI**
  - 1464 + 1467
  - 5065 + 5785

- **WTCCC/UKT2D**
  - 1924 + 2938
  - 3757 + 5346

- **Total**
  - 4549 + 5579
  - 10037 + 12389
<table>
<thead>
<tr>
<th>Gene(s)</th>
<th>FUSION</th>
<th>DGI</th>
<th>UK</th>
<th>All Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR</td>
<td>OR</td>
<td>OR</td>
<td>OR</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>p-value</td>
<td>p-value</td>
<td>p-value</td>
</tr>
<tr>
<td>TCF7L2</td>
<td>1.34</td>
<td>1.38</td>
<td>1.37</td>
<td>1.37</td>
</tr>
<tr>
<td></td>
<td>1.3 x 10^{-8}</td>
<td>2.3 x 10^{-31}</td>
<td>6.7 x 10^{-13}</td>
<td>1.0 x 10^{-48}</td>
</tr>
<tr>
<td>CDKN2A/B</td>
<td>1.20</td>
<td>1.20</td>
<td>1.19</td>
<td>1.20</td>
</tr>
<tr>
<td></td>
<td>.0022</td>
<td>5.4 x 10^{-8}</td>
<td>4.9 x 10^{-7}</td>
<td>7.8 x 10^{-15}</td>
</tr>
<tr>
<td>IGF2BP2</td>
<td>1.18</td>
<td>1.17</td>
<td>1.11</td>
<td>1.14</td>
</tr>
<tr>
<td></td>
<td>2.1 x 10^{-4}</td>
<td>1.7 x 10^{-9}</td>
<td>1.6 x 10^{-4}</td>
<td>8.9 x 10^{-16}</td>
</tr>
<tr>
<td>FTO</td>
<td>1.11</td>
<td>1.03</td>
<td>1.23</td>
<td>1.17</td>
</tr>
<tr>
<td></td>
<td>.016</td>
<td>.25</td>
<td>7.3 x 10^{-14}</td>
<td>1.3 x 10^{-12}</td>
</tr>
<tr>
<td>CDKAL1</td>
<td>1.12</td>
<td>1.08</td>
<td>1.16</td>
<td>1.12</td>
</tr>
<tr>
<td></td>
<td>.0095</td>
<td>.0024</td>
<td>1.3 x 10^{-8}</td>
<td>4.1 x 10^{-11}</td>
</tr>
<tr>
<td>KCNJ11</td>
<td>1.11</td>
<td>1.15</td>
<td>1.15</td>
<td>1.14</td>
</tr>
<tr>
<td></td>
<td>.013</td>
<td>1.0 x 10^{-7}</td>
<td>.0013</td>
<td>6.7 x 10^{-11}</td>
</tr>
<tr>
<td>HHEX, IDE</td>
<td>1.10</td>
<td>1.14</td>
<td>1.13</td>
<td>1.13</td>
</tr>
<tr>
<td></td>
<td>.026</td>
<td>1.7 x 10^{-4}</td>
<td>4.6 x 10^{-6}</td>
<td>5.7 x 10^{-10}</td>
</tr>
<tr>
<td>SLC30A8</td>
<td>1.18</td>
<td>1.07</td>
<td>1.12</td>
<td>1.12</td>
</tr>
<tr>
<td></td>
<td>7.0 x 10^{-5}</td>
<td>.047</td>
<td>7.0 x 10^{-5}</td>
<td>5.3 x 10^{-8}</td>
</tr>
<tr>
<td>Chr 11</td>
<td>1.48</td>
<td>1.16</td>
<td>1.13</td>
<td>1.25</td>
</tr>
<tr>
<td></td>
<td>5.7 x 10^{-8}</td>
<td>.12</td>
<td>.068</td>
<td>4.3 x 10^{-7}</td>
</tr>
<tr>
<td>PPARG</td>
<td>1.20</td>
<td>1.09</td>
<td>1.23</td>
<td>1.14</td>
</tr>
<tr>
<td></td>
<td>.0014</td>
<td>.019</td>
<td>.0013</td>
<td>1.7 x 10^{-6}</td>
</tr>
</tbody>
</table>
Arg325Trp Variant in \textit{SLC30A8}

- Non-synonymous variant in zinc transporter specific to pancreatic beta-cell
- \textit{SLC30A8} transports zinc from cytoplasm into insulin secretory vesicles, where insulin stored as hexamer bound with two Zn$^{++}$ ions prior to secretion
- May affect zinc accumulation in insulin granules, affecting stability, storage, or secretion
Chr 11 gene desert
Comparison to French/Canadian GWA

FUSION / DGI / WTCCC-UKT2D confirm top three loci (*TCF7L2, SLC30A8, HHEX*)

No support for other loci, although rs9300039 within 0.3 and 2.4 Mb of Sladek et al. chromosome 11 regions
deCODE T2D GWA

- GWA of 1399 T2D cases, 5275 controls, all from Iceland genotyped for Illumina 317K chip
- 47 SNPs followed up in Danish sample of 1110 cases and 2272 controls
- Subsequent follow up in several additional samples
- Evidence for association with variants in *TCF7L2*, *CDKAL1*, *SLC30A8*
- Five more T2D GWAs subsequently published
Ten Loci for T2D: Comments

- Of seven new loci, only one (*HHEX, IDE*) included in our prior list of >200 candidate genes

- **SLC30A8** locus: non-synonymous SNP in excellent candidate gene

- For other new loci, SNPs intronic (e.g. *IGF2BP2, CDKAL1*) or just near genes; likely not actual risk variants

- Chr 11 locus >1 Mb from nearest annotated gene
Cross-Study Analyses Including CAD, Obesity

- *FTO* result appears to be mediated primarily through obesity (Frayling et al. 2007, Dina et al. 2007)

- *CDKN2A/B* region SNPs identified in GWA of myocardial infarction (McPherson et al. 2007, Helgadottir et al. 2007)
  - cyclin dependent kinase inhibitors implicated in various cancers
Current T2D GWA Meta-Analysis

- Imputation in UK (Impute), DGI (Mach) samples
- Meta-analysis of 2.3 million genotyped or imputed SNPs
- Chose 58 best SNPs for genotyping in GWA and follow up samples (total N~35,000)
- New signals: 10 with p<10^{-6}, 5 with p<10^{-7}, 2 with p<10^{-8}
- Paper in preparation, presentation next week at ASHG (L Scott et al.)
GWAs of T2D-Related QTs

• Once genotyping completed, GWAs for other traits “free”

• Pursuing glucose/insulin, anthropometrics, lipids, blood pressure

• Many samples potentially available: GWA, follow up
  – primary GWA sharing with DGI, SardiNIA
  – follow up with many groups
  – organizationally complex

• Clear evidence for glucose locus, height locus, ≥15 lipid loci (≥5 novel); more soon
## Best Novel Lipid Meta-Analysis Results

<table>
<thead>
<tr>
<th>Trait</th>
<th>Chr</th>
<th>Pos (Mb)</th>
<th>Effect (mg/dl)</th>
<th>P-value</th>
<th>Nearby Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>HDL</td>
<td>1</td>
<td>226.6</td>
<td>0.72</td>
<td>$1 \times 10^{-8}$</td>
<td>GALNT2</td>
</tr>
<tr>
<td>HDL</td>
<td>12</td>
<td>108.4</td>
<td>0.56</td>
<td>$2 \times 10^{-8}$</td>
<td>MVK, MMAB</td>
</tr>
<tr>
<td>LDL</td>
<td>1</td>
<td>109.5</td>
<td>4.45</td>
<td>$6 \times 10^{-22}$</td>
<td>CELSR2, PSRC1, SORT1</td>
</tr>
<tr>
<td>LDL</td>
<td>19</td>
<td>19.5</td>
<td>5.97</td>
<td>$6 \times 10^{-12}$</td>
<td>NCAN, CILP2</td>
</tr>
<tr>
<td>TG</td>
<td>19</td>
<td>19.5</td>
<td>7.48</td>
<td>$3 \times 10^{-9}$</td>
<td>NCAN, CILP2</td>
</tr>
<tr>
<td>TG</td>
<td>8</td>
<td>126.6</td>
<td>7.27</td>
<td>$8 \times 10^{-13}$</td>
<td>TRIB1</td>
</tr>
</tbody>
</table>
Summary and Comments (1)

- Comprehensive GWAs feasible, ≥10 reported to date for T2D in last six months
- Identified/confirmed 10 common variants associated with T2D risk; all of modest effect, all could lead to drug targets
- Identified apparently non-synonymous risk variant in SLC30A8
Summary and Comments (2)

- Joint analysis of multiple T2D studies likely needed to identify additional T2D risk variants; in process (up to 10 new loci)

- Progress even for “geneticist’s nightmare”

- Parallel GWAs identify novel loci for glucose, height, lipids (5), plus several common variants in known lipid loci
FUSION and CIDR

**UNC-Chapel Hill**
Karen Mohlke
Kyle Gaulton
Jason Luo
Li Qin

**NHGRI / NIH**
Francis Collins
Lori Bonnycastle
Peter Chines
Michael Erdos
Narisu Narisu
L. Prokunina
Nancy Riebow
Andrew Sprau
Amy Swift
Maurine Tong

**U Michigan**
Karen Conneely
Charles Ding
William Duren
Terry Gliedt
Kevin He
Larry Hu
Anne Jackson
Laura Scott
Heather Stringham
Peggy White
Cristen Willer
Fang Xiang
Rui Xiao

**National Public Health Institute Helsinki**
Jaakko Tuomilehto
Timo Valle

**USC**
Richard Bergman
Thomas Buchanan
Richard Watanabe

**CIDR**
Kimberly Doheny
Elizabeth Pugh
and many others

**Calvin College**
Randall Pruim
T2D Collaborating Groups

**Diabetes Genetics Initiative (DGI)**  
Broad Institute, Lund University, Novartis

David Altshuler  
Thomas Hughes

Peter Almgren  
Paul de Bakker  
Brendan Blumenstiel  
Noël Burtt  
Hong Chen  
Mark Daly  
Jose Florez  
Stacey Gabriel  
Candace Guiducci

Leif Groop

Joel Hirschhorn  
Sekar Kathiresan  
Valeriya Lyssenko  
Joanne Meyer  
Jeffrey Roix  
Richa Saxena  
Benjamin Voigt

**Wellcome Trust Case Control Consortium (WTCCC) and UK T2D Genetics Consortium**

Mark McCarthy

Jeffrey Barrett  
Lon Cardon  
Alex Doney  
Peter Donnelly  
Sian Ellard  
Katherine Elliott  
Timothy Frayling  
Rachel Freathy  
Christopher Groves  
Graham Hitman  
Lorna Harries  
Beatrice Knight  
Hana Lango

Andrew Hattersley

Cecilia Lindgren  
Jonathon Marchini  
Andrew Morris  
Katharine Owen  
Colin Palmer  
John Perry  
Nigel Rayner  
Beverly Shields  
Nicholas Timpson  
Mark Walker  
Michael Weedon  
Eleftheria Zeggini
Lipids Collaborating Groups

- SardiNIA: David Schlessinger, Gonçalo Abecasis, Serena Sanna, Angelo Scuteri, Samer Najjar, James Strait, Andrea Maschio, Fabio Busonero, Giuseppe Albai, Wei-Min Chen, Ramaiah Nagaraja, Manuela Uda, Antonio Cao, Ed Lakatta

- DGI

- UK: Robert Clarke, Derrick Bennett, Sarah Parish, Rory Collins

- France: Mark Lathrop, Simon Heath, Pilar Galan, Pierre Meneton, Serge Herçberg, Diana Zelenika

- Maryland: Alan Shuldiner, Haiqing Shen