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

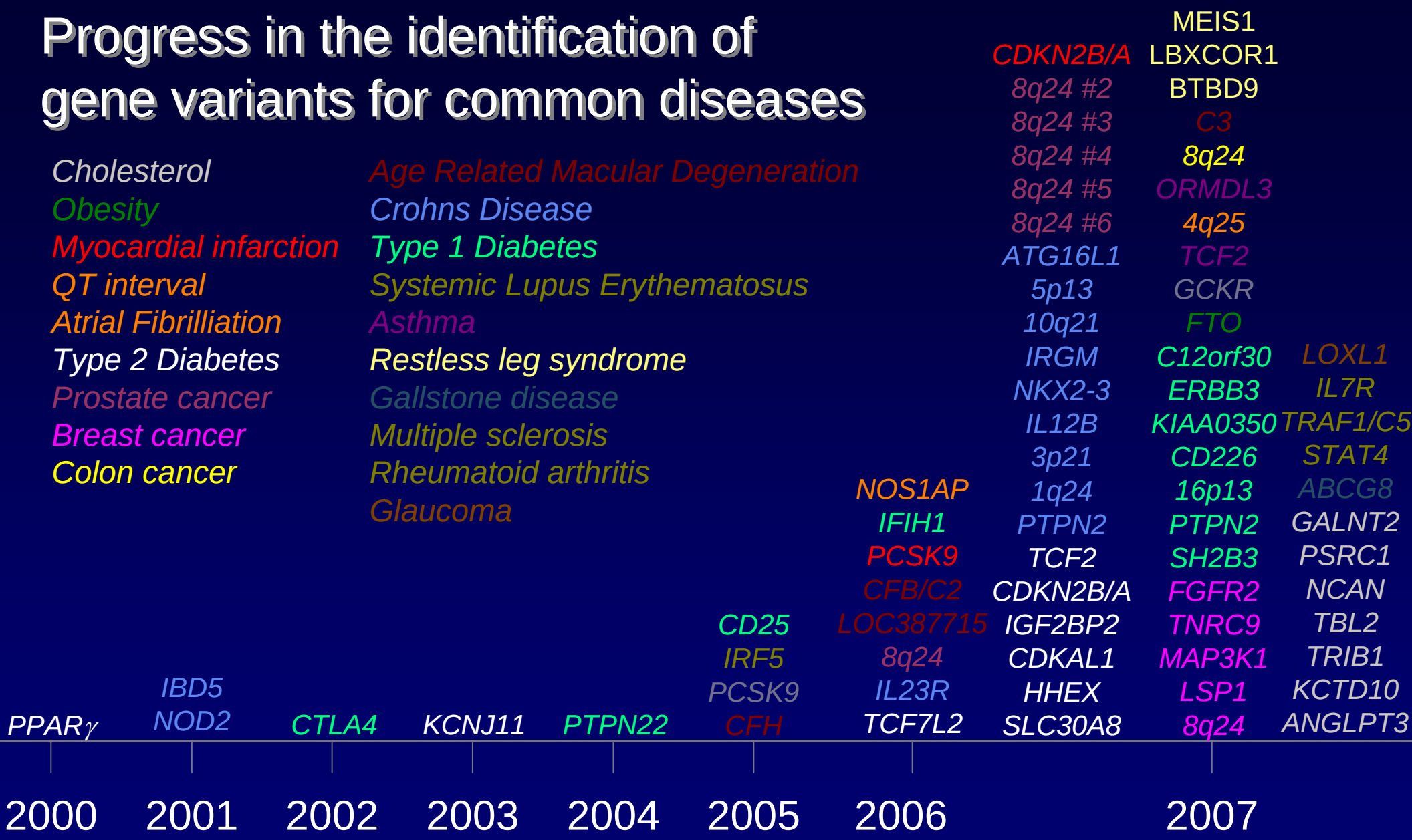
Introduction

- 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

Cholesterol
Obesity
Myocardial infarction
QT interval
Atrial Fibrilliation
Type 2 Diabetes
Prostate cancer
Breast cancer
Colon cancer

Age Related Macular Degeneration
Crohns Disease
Type 1 Diabetes
Systemic Lupus Erythematosus
Asthma
Restless leg syndrome
Gallstone disease
Multiple sclerosis
Rheumatoid arthritis
Glaucoma



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, WTCCC
- 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

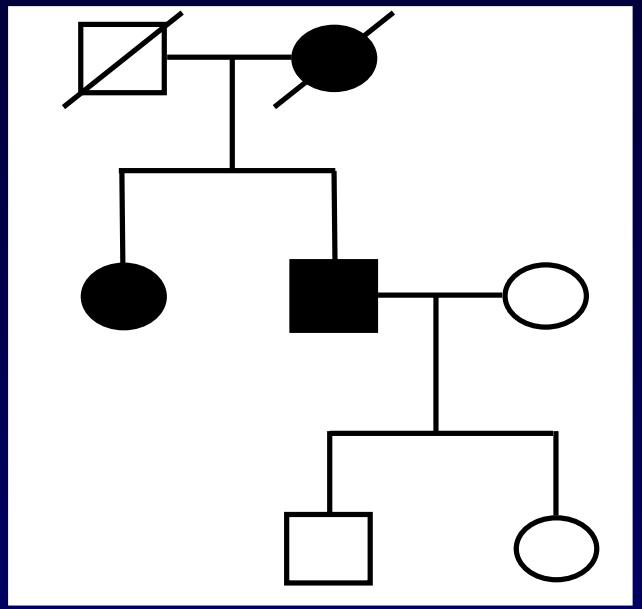
Identify genetic variants that predispose to type 2 diabetes (T2D) or are responsible for variability in T2D-related quantitative traits (QTs)

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^{o} 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)

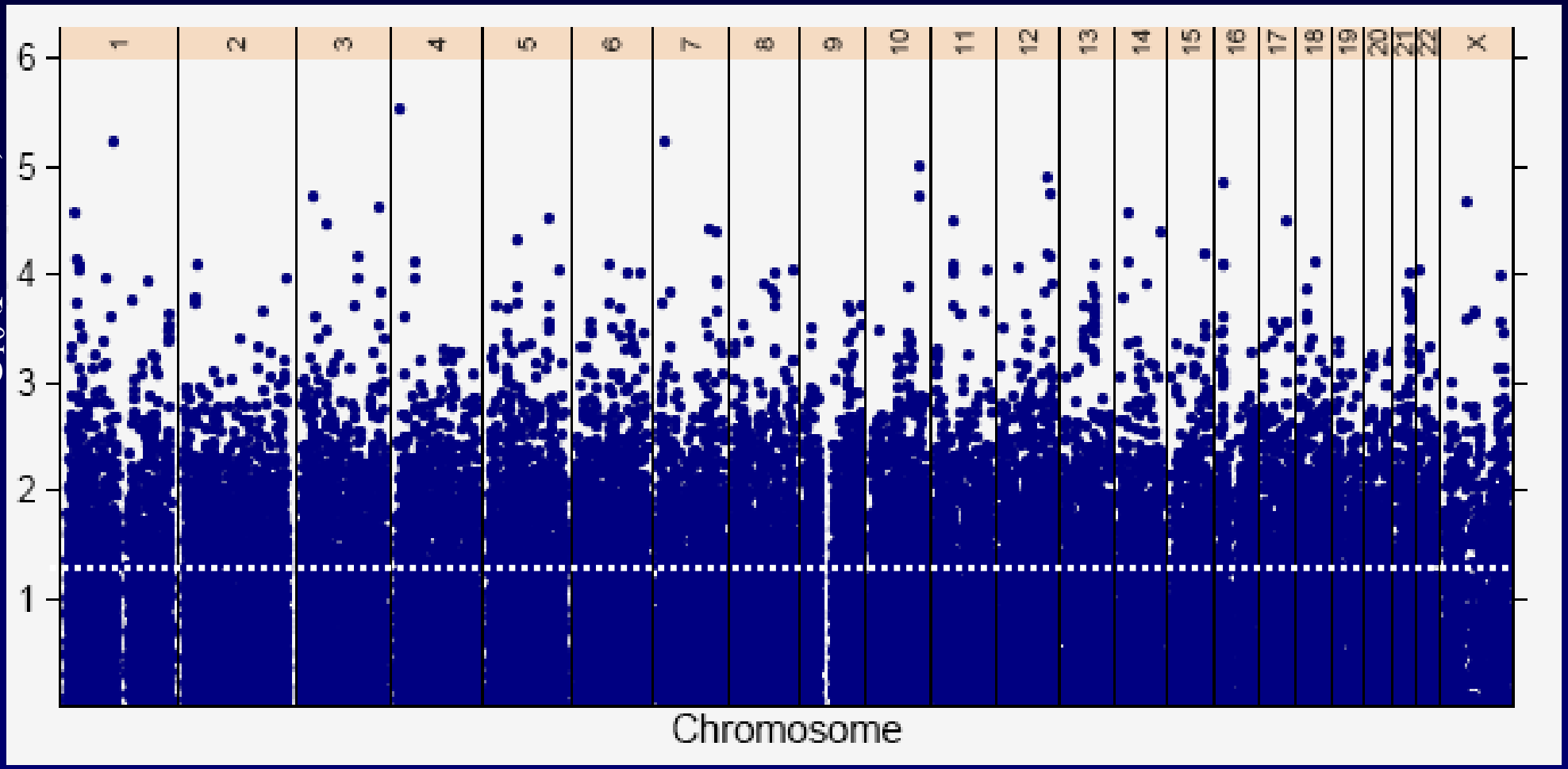
- 1,808 SNPs (0.6%) excluded from analysis
  - Hardy-Weinberg equilibrium p-value < 10^{-6}
  - < 90% successful genotypes
  - > 3 Mendelian or duplicate errors
  - < 10 minor alleles
- 4,881 SNPs (1.5%) flagged for analysis
  - Hardy-Weinberg equilibrium p-value < 10^{-4}
  - < 95% successful genotypes
  - >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

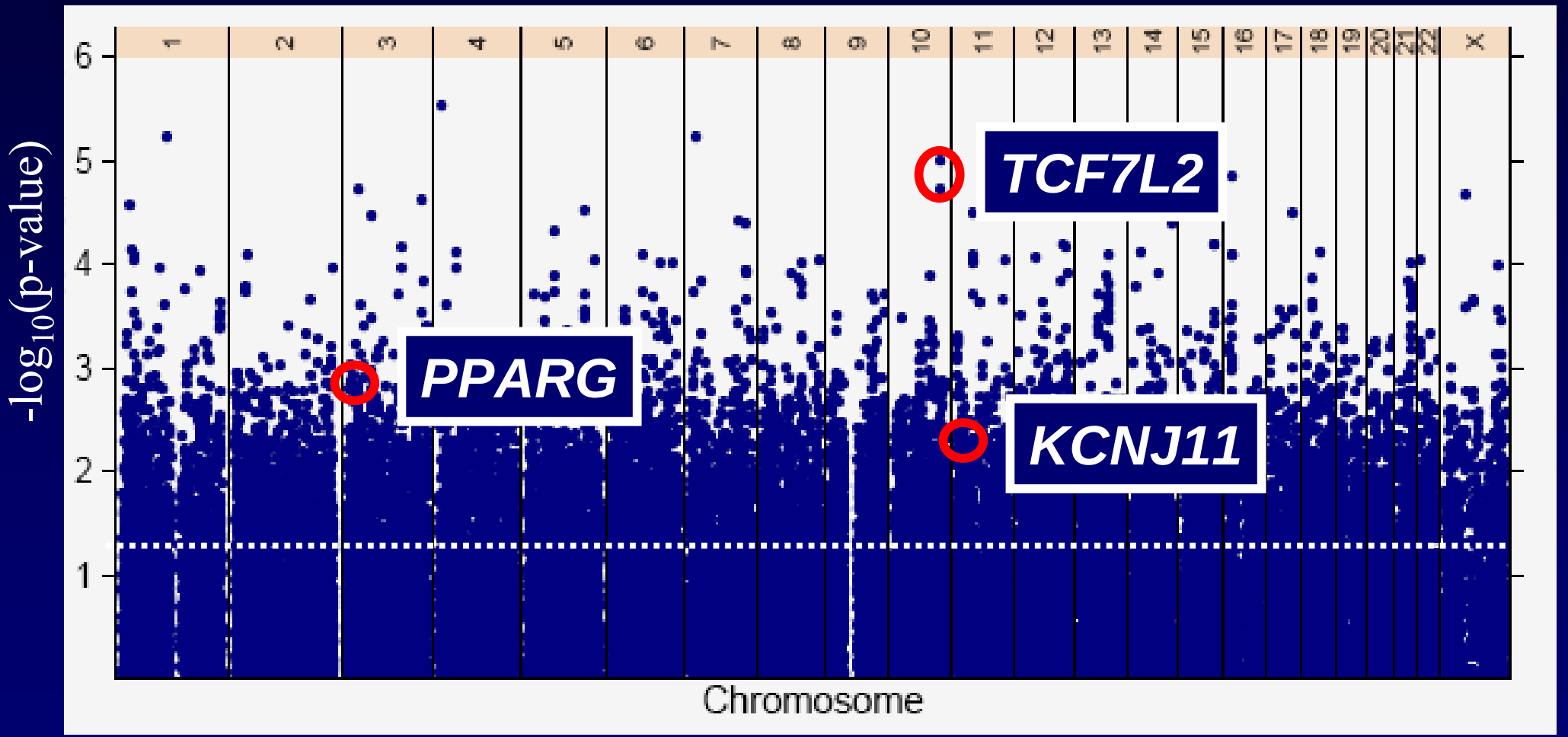
1161 Finnish T2D cases + 1174 Finnish NGT controls



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?

Threshold	Expected Number	Observed Number	Empirical p-value
p-value < 10^{-6}	0.3	0	1
p-value < 10^{-5}	3	3	.54
p-value < 10^{-4}	31	43	.19

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%

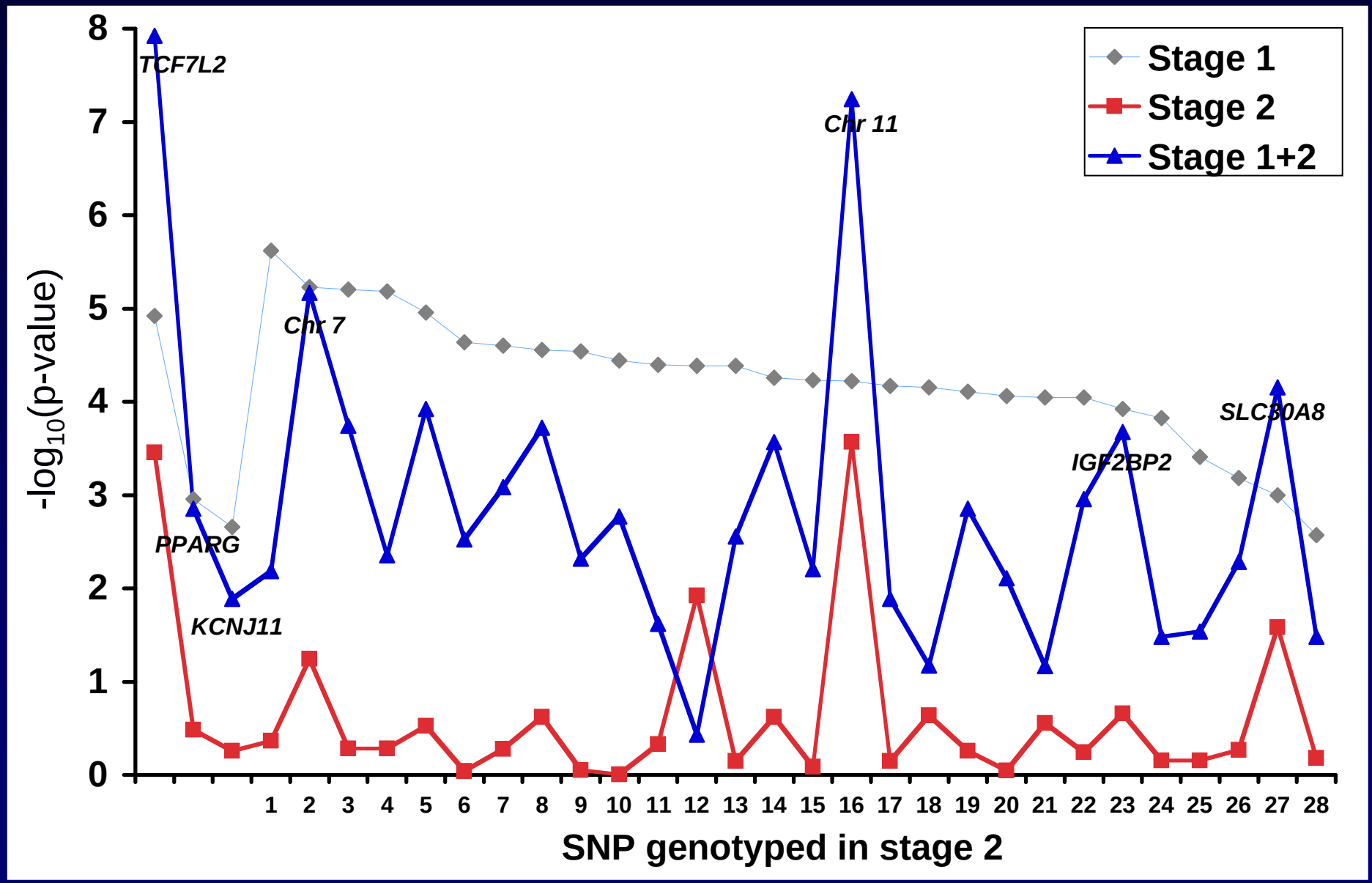
Imputed vs. Genotyped SNP Results

Allele frequency		P-value		Odds ratio	
Imputed	Genotyped	Imputed	Genotyped	Imputed	Genotyped
.024	.021	2.5×10^{-6}	6.3×10^{-6}	2.57	2.20
.543	.540	5.3×10^{-6}	1.1×10^{-5}	1.33	1.31
.114	.136	2.0×10^{-5}	4.1×10^{-5}	1.47	1.41
.494	.490	6.6×10^{-5}	5.5×10^{-5}	1.28	1.28
.927	.924	7.5×10^{-5}	9.0×10^{-5}	1.72	1.65
.744	.753	1.4×10^{-4}	3.9×10^{-4}	1.33	1.30
.289	.291	1.7×10^{-4}	1.2×10^{-4}	1.27	1.28
.970	.973	1.9×10^{-4}	3.6×10^{-5}	2.47	2.58
.401	.361	6.3×10^{-4}	1.6×10^{-3}	1.26	1.22
.817	.816	9.5×10^{-4}	1.0×10^{-3}	1.31	1.30
.605	.605	9.9×10^{-4}	1.2×10^{-3}	1.23	1.22

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



Meta-Analysis of Three T2D GWAs

- 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

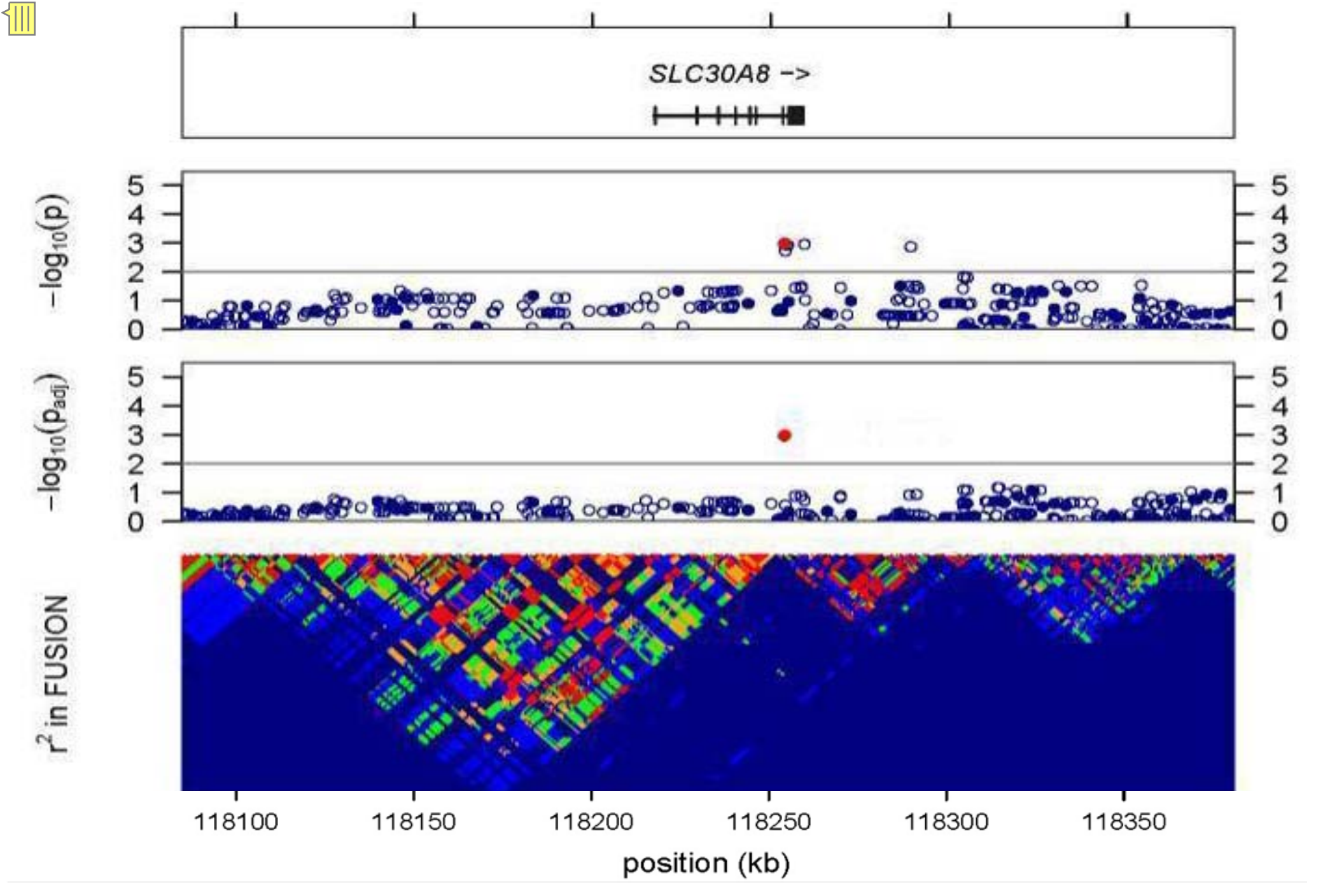
 # cases + # controls

- **FUSION**
  - **1161 + 1174**
  - **1215 + 1258**
- **DGI**
  - **1464 + 1467**
  - **5065 + 5785**
- **WTCCC/UKT2D**
  - **1924 + 2938**
  - **3757 + 5346**
- **Total**
  - **4549 + 5579**
  - **10037 + 12389**

Top Results of T2D GWA Meta-Analysis

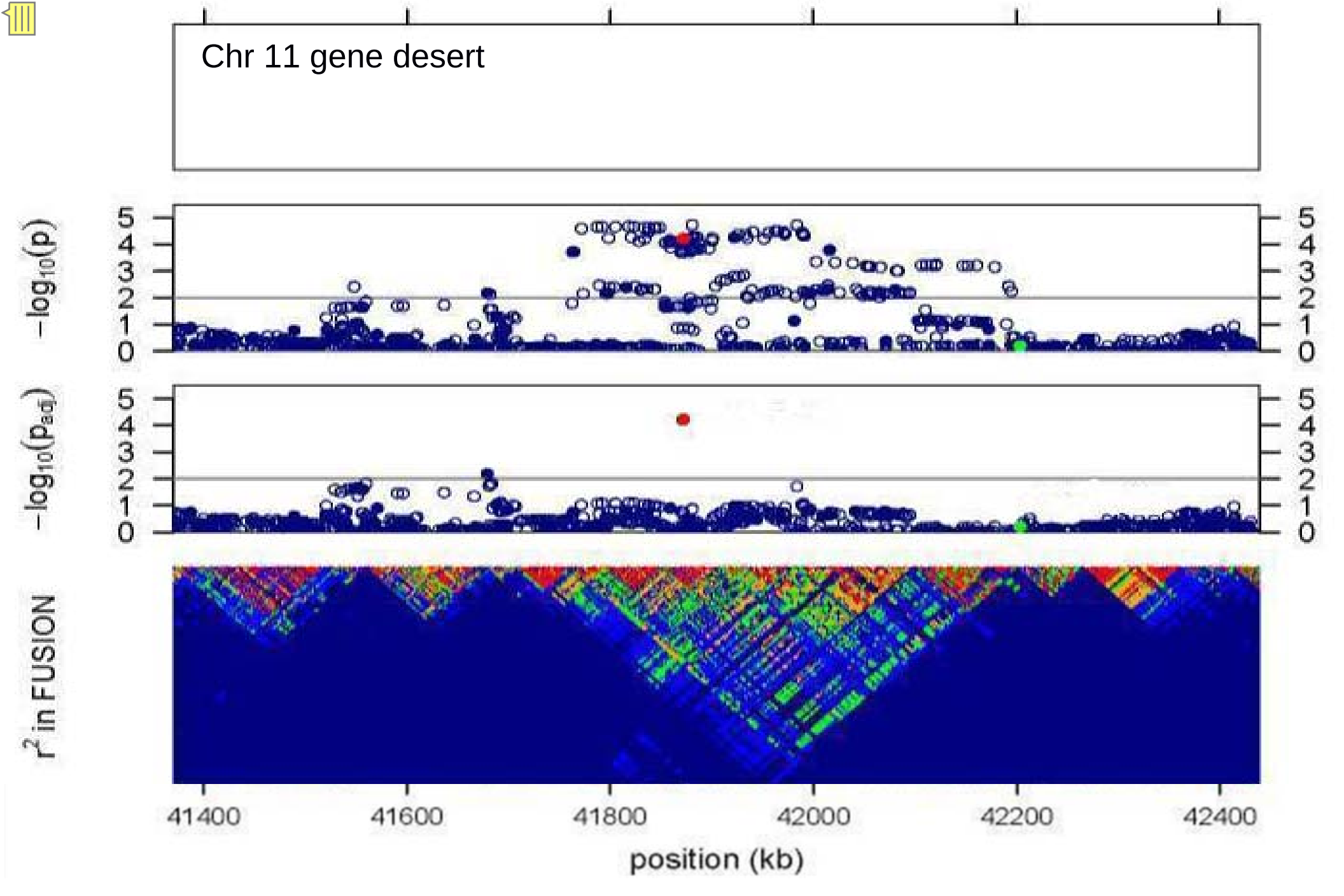
Scott et al. *Science* June 2007: WTCCC/UKT2D, DGI, FUSION

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


SLC30A8 ->
−log10(p)
−log10(padj)
r2 in FUSION
0
1
2
3
4
5
118100
118150
118200
118250
118300
118350
position (kb)

Arg325Trp Variant in *SLC30A8*

- Non-synonymous variant in zinc transporter specific to pancreatic beta-cell
- 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
−log10(p)
−log10(padj)
r2 in FUSION
0
1
2
3
4
5
41400
41600
41800
42000
42200
42400
position (kb)

Comparison to French/Canadian GWA

A genome-wide association study identifies novel risk loci for type 2 diabetes

Robert Sladek[1,2,4], Ghislain Rocheleau[1*], Johan Rung[4*], Christian Dina[5*], Lishuang Shen[1], David Serre[1], Philippe Boutin[5], Daniel Vincent[4], Alexandre Belisle[4], Samy Hadjadj[6], Beverley Balkau[7], Barbara Heude[7], Guillaume Charpentier[8], Thomas J. Hudson[4,9], Alexandre Montpetit[4], Alexey V. Pshezhetsky[10], Marc Prentki[10,11], Barry I. Posner[2,12], David J. Balding[13], David Meyre[5], Constantin Polychronakos[1,3] & Philippe Froguel[5,14]

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

Trait	Chr	Pos (Mb)	Effect (mg/dl)	P-value	Nearby Genes
HDL	1	226.6	0.72	1×10^{-8}	*GALNT2*
HDL	12	108.4	0.56	2×10^{-8}	*MVK, MMAB*
LDL	1	109.5	4.45	6×10^{-22}	*CELSR2, PSRC1, SORT1*
LDL	19	19.5	5.97	6×10^{-12}	*NCAN, CILP2*
TG	19	19.5	7.48	3×10^{-9}	*NCAN, CILP2*
TG	8	126.6	7.27	8×10^{-13}	*TRIB1*

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

Calvin College

Randall Pruim

U Michigan

Gonçalo Abecasis
Yun Li
Jun Ding
Paul Scheet

CIDR

Kimberly Doheny
Elizabeth Pugh
and many others

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