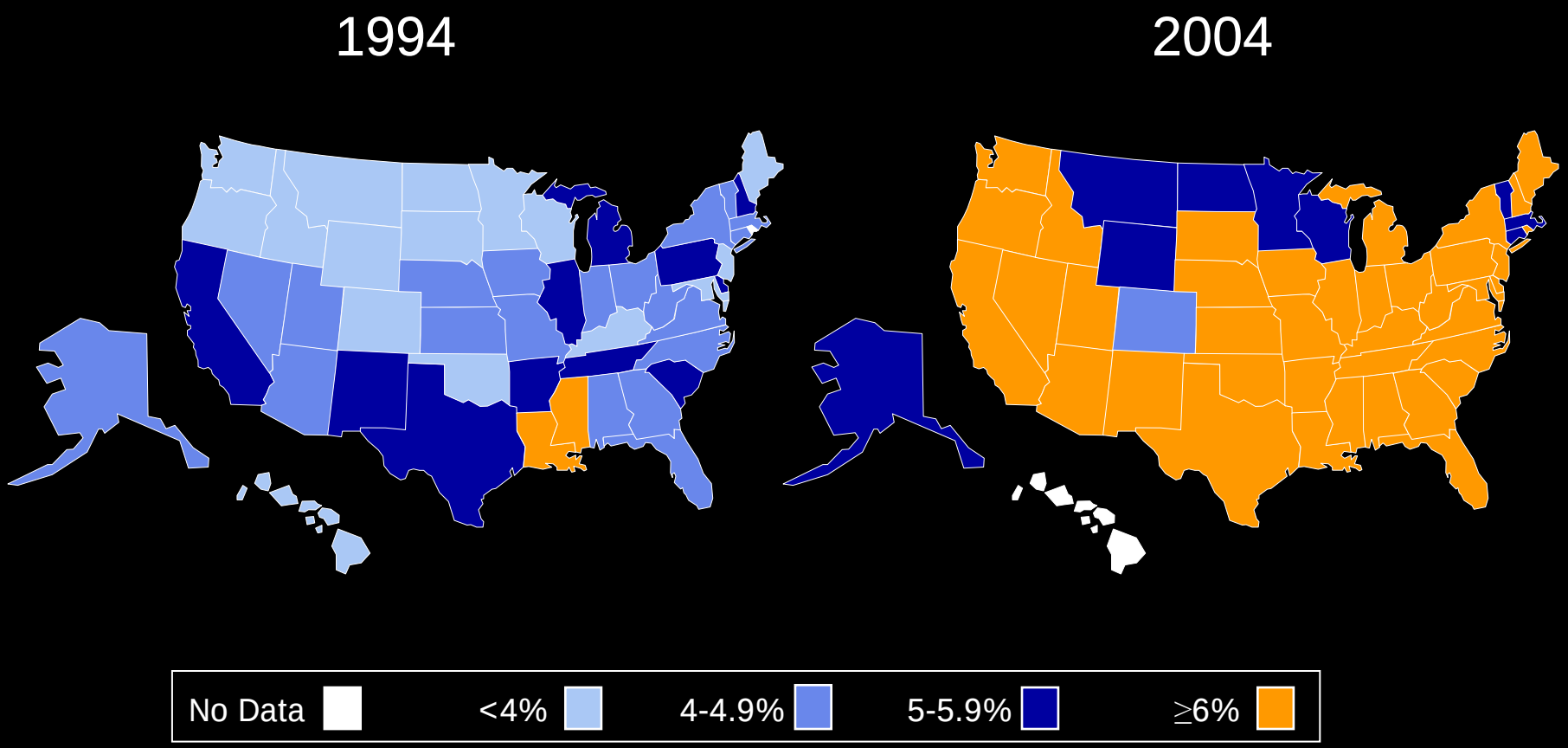
Dissemination/Publication: Case Study on Type 2 Diabetes

Multi-IC Symposium on GWAS

Francis S. Collins, M.D., Ph.D.
National Human Genome Research Institute
May 22, 2007

Estimates of Diagnosed Diabetes Among Adults in the U.S.



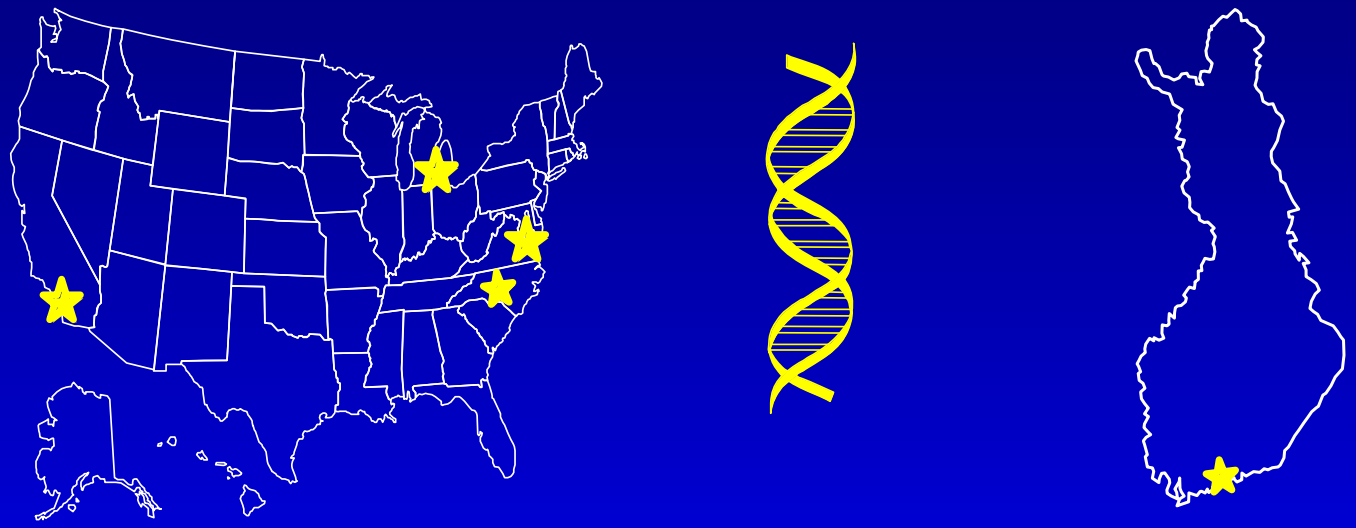
CDC
Centers for Disease Control and Prevention

Type 2 Diabetes: "The geneticist's nightmare"

- **Family history as a substantial risk factor**
  - **Relative risk to a sibling ~3.5**
- **Environment as a major contributor**
- **Family linkage studies relatively disappointing**
- **Validated genes prior to 2007:**
  - ***PPARG (candidate gene)***
  - ***KCNJ11 (candidate gene)***
  - ***TCF7L2 (linkage study)***

The FUSION Study

Finland-United States Investigation of NIDDM Genetics



Subject Recruitment and Clinical Testing

National Public Health Institute, Helsinki, Finland

Molecular Genetics

National Human Genome Research Institute, Bethesda, MD

University of North Carolina, Chapel Hill, NC

Biochemical Measurements

USC Keck School of Medicine, Los Angeles, CA

Statistical Analysis

University of Michigan School of Public Health, Ann Arbor, MI

Applying Genome Wide Association to Type 2 Diabetes

- Three groups, each with 1000 – 1500 cases and 1000 – 3000 controls in Stage 1:
  - FUSION (Boehnke, Bergman, Collins, Mohlke, Tuomilehto)
  - Diabetes Genetics Initiative of Broad, Novartis, and Lund (Altshuler, Groop)
  - Wellcome Trust Case Control Consortium/UK Type 2 Diabetes Consortium (McCarthy, Hattersley, Donnelly)
- Genotyped with Illumina 317K or Affy 500K panel
- Compared results across all three studies
- Followed up promising signals in Stage 2 validation set

Three Groups Working Together

 # cases + controls

FUSION

S1: 1161 + 1174

S2: 1215 + 1258

DGI

S1: 1464 + 1467

S2: 5065 + 5785

WTCCC/UKT2D

S1: 1924 + 2938

S2: 3757 + 5346

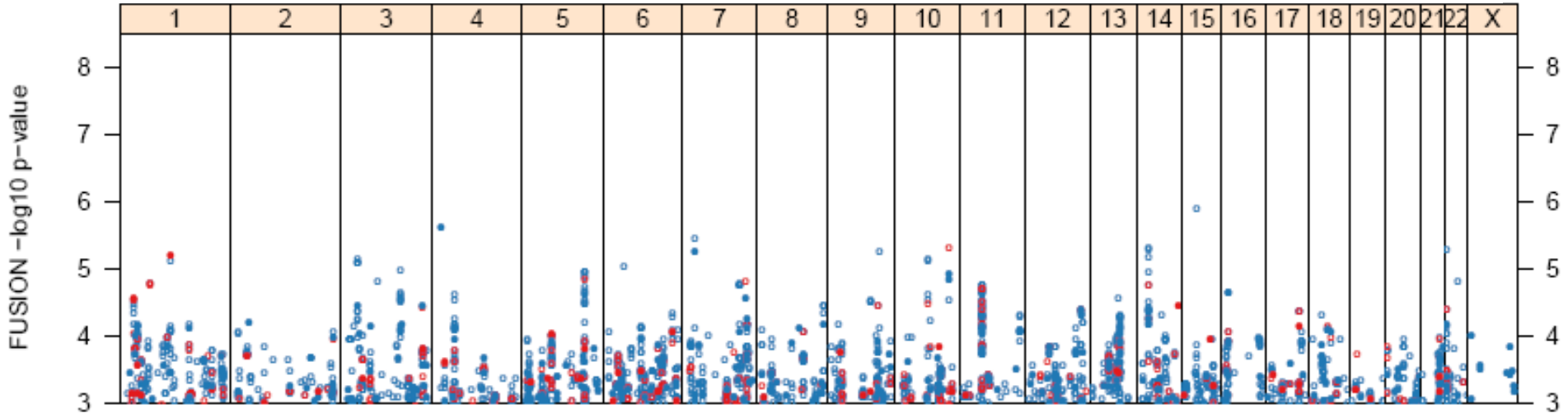
Totals

S1 = 4549 + 5579

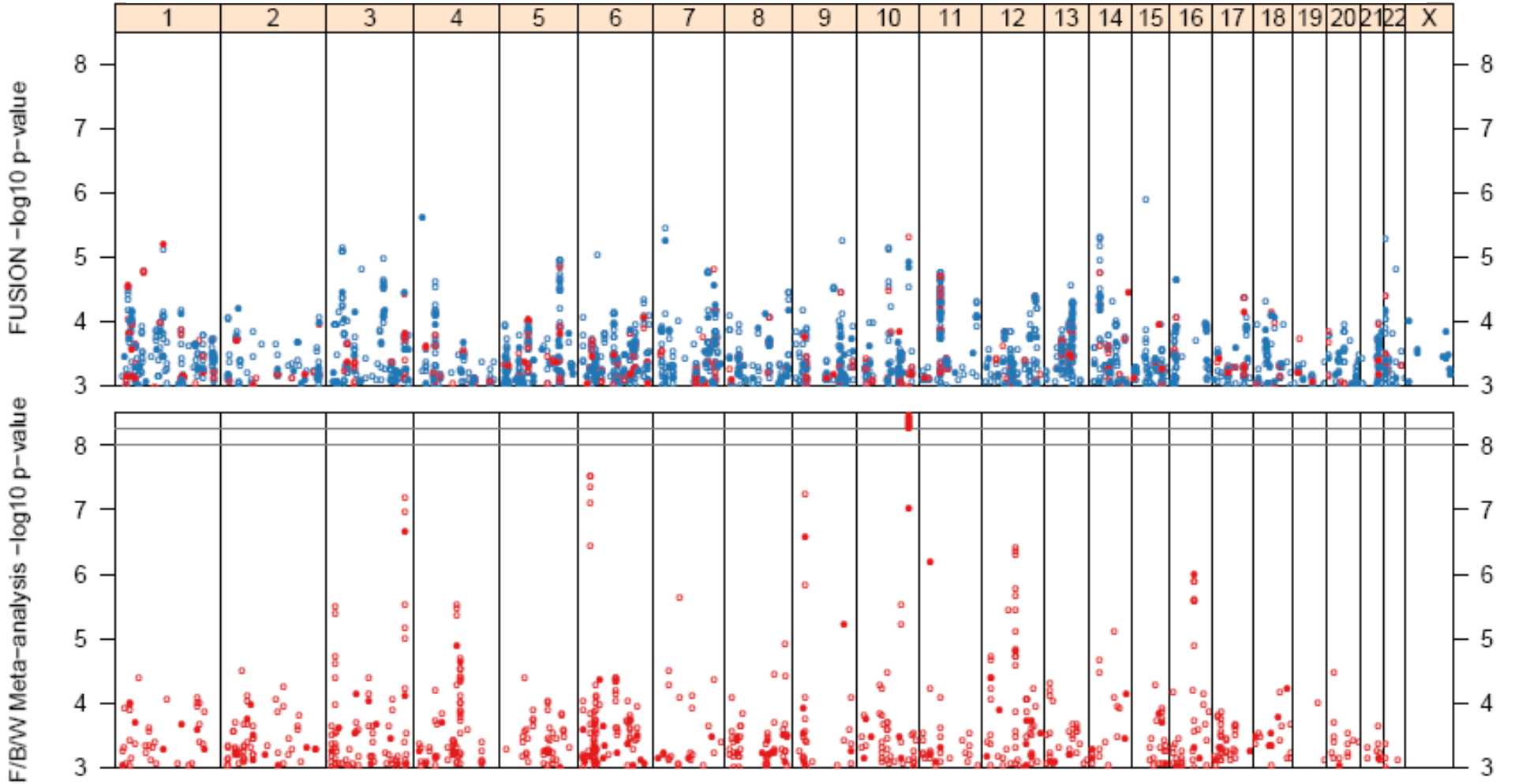
S2 = 10053 + 12389

(n=32,554)

Stage 1: FUSION only (1161 cases + 1174 controls)



Stage 1 – FUSION only

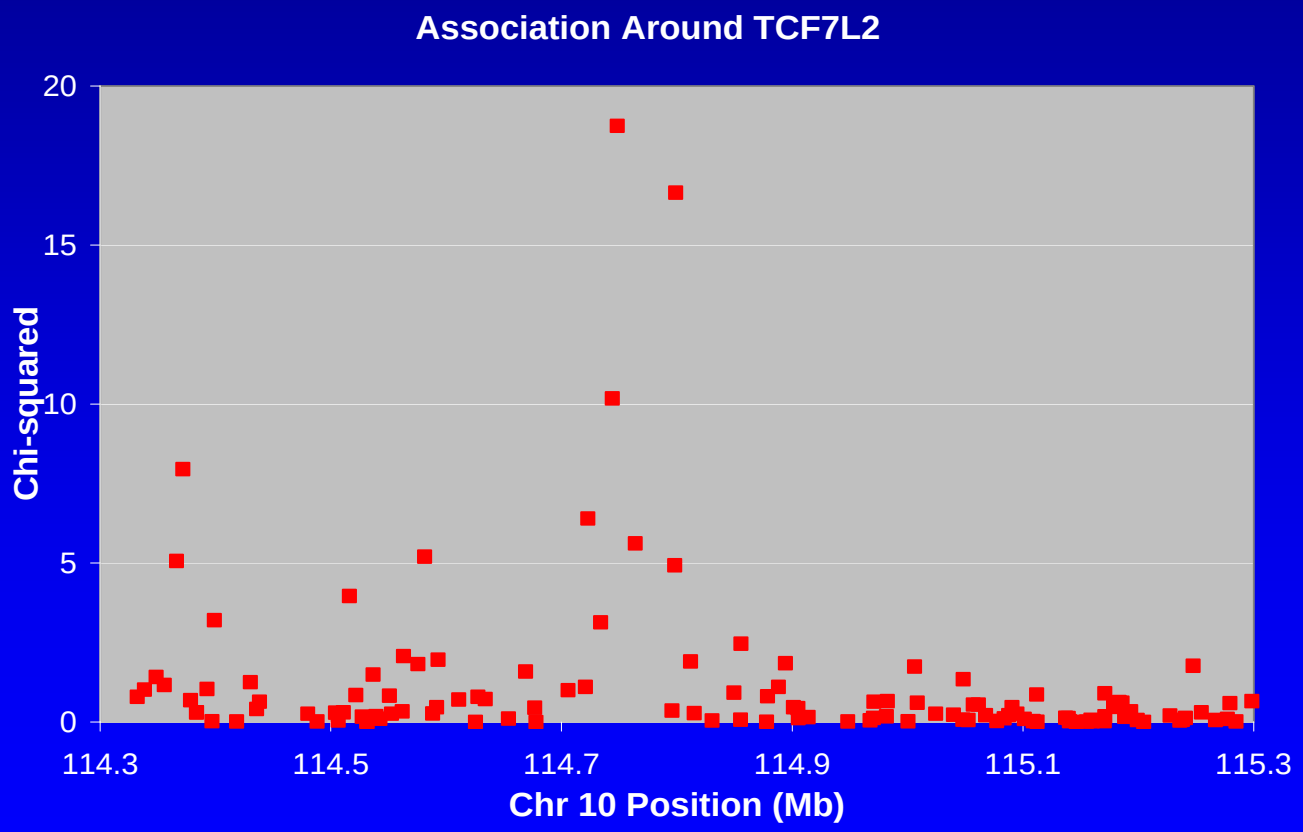


Stage 1 – FUSION + DGI + WTCCC
(4549 cases + 5579 controls)

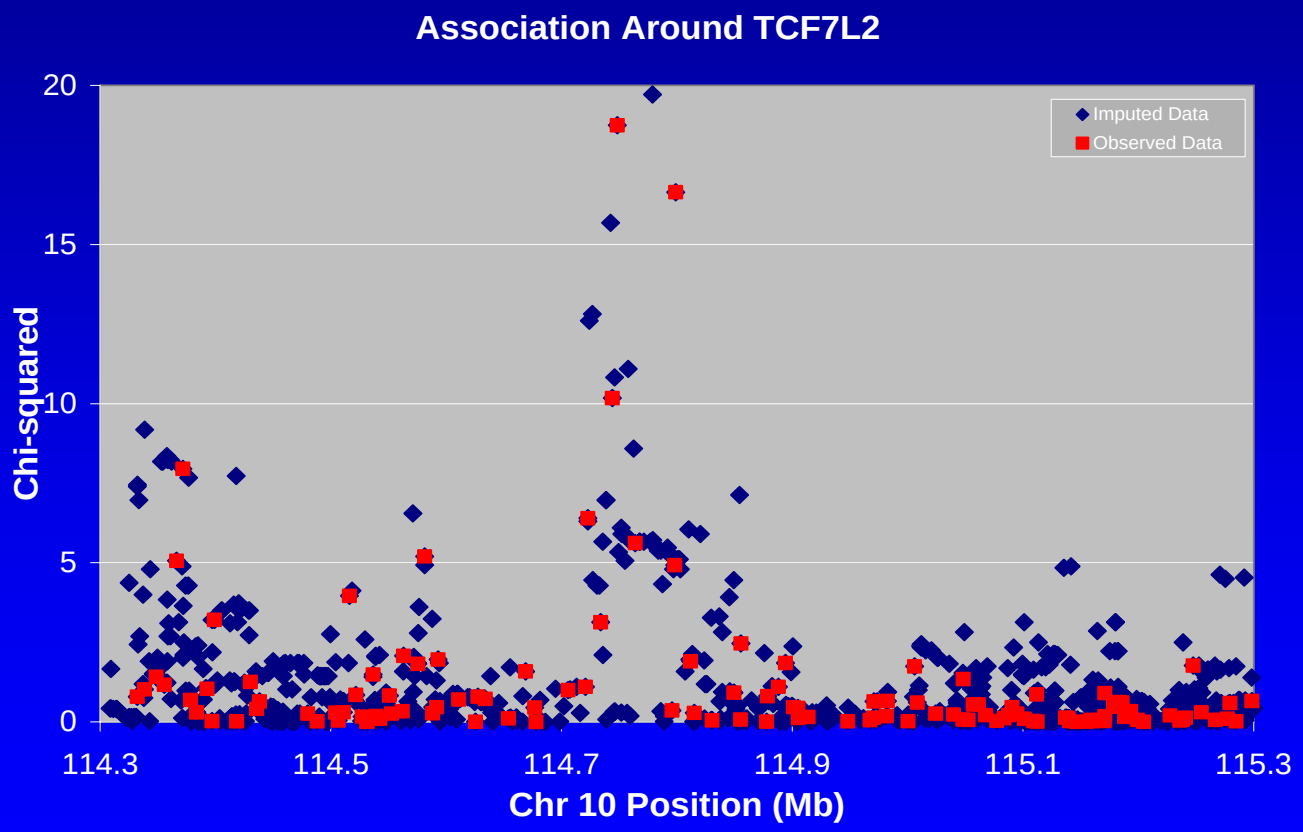
Imputing Missing Genotypes in Case Control Samples

- Methods and software have now been developed and tested by
  - Goncalo Abecasis, Michigan
  - Jonathan Marchini, Oxford
- Begins with GWA data from panel of choice
- Uses HapMap data from similar geographic origins to infer what alleles were most likely present at untyped loci
- Limited to SNPs in strong LD with typed SNPs
- Can produce quality score estimates
- Allows merging of data sets from Illumina, Affymetrix, or Perlegen panels

What We'd Like To Do:
Start with Measured Genotypes



What We'd Like To Do: Estimate Effects of Untyped Markers…



Does This Actually Work?

- Used about ~300,000 SNPs from Illumina HumanHap300 to impute 2.1 million HapMap SNPs in ~2500 individuals from the FUSION study of type 2 diabetes

- Compared imputed genotypes with actual experimental genotypes in a candidate region on chromosome 14
  - 1190 individuals, 521 markers not on Illumina HumanHap300

- Results of comparison
  - Average r^2 with true genotypes 0.92
  - 1.4% of imputed alleles mismatch original
  - Most errors concentrated on worst 3% of SNPs

Top 10 Results From Combined Analysis Stage 1 + Stage 2, n = 32,554

	FUSION		DGI		WTCCC/UKT2D		All Samples	
Gene	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}
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}
CDKN2A/B	1.20	.0022	1.20	5.4×10^{-8}	1.19	4.9×10^{-7}	1.20	7.8×10^{-15}
FTO	1.11	0.016	1.03	0.25	1.23	7.3×10^{-14}	1.17	1.3×10^{-12}
CDKAL1	1.12	0.0095	1.08	0.0024	1.16	1.3×10^{-8}	1.12	4.1×10^{-11}
KCNJ11	1.11	0.013	1.15	1.0×10^{-7}	1.15	0.0013	1.14	6.7×10^{-11}
HHEX	1.10	0.026	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	0.047	1.12	7.0×10^{-5}	1.12	5.3×10^{-8}
Chr 11	1.48	5.7×10^{-8}	1.16	0.12	1.13	0.068	1.23	4.3×10^{-7}
PPARG	1.20	0.0014	1.09	0.019	1.23	0.0013	1.14	1.7×10^{-6}

Sciencexpress Report

A Genome-Wide Association Study of Type 2 Diabetes in Finns Detects Multiple Susceptibility Variants

Laura J. Scott,[1] Karen L. Mohlke,[2] Lori L. Bonnycastle,[3] Cristen J. Willer,[1] Yun Li,[1] William L. Duren,[1] Michael R. Erdos,[3] Heather M. Stringham,[1] Peter S. Chines,[3] Anne U. Jackson,[1] Ludmila Prokunina-

Sciencexpress Report

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

Sciencexpress Report

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

Eleftheria Zeggini,[1,2*] Michael N. Weedon,[3,4*] Cecilia M. Lindgren,[1,2*] Timothy M. Frayling,[3,4*] Katherine S. Elliott,[2] Hana Lango,[3,4] Nicholas J. Timpson,[2,5] John R. B. Perry,[3,4] Nigel W. Rayner,[1,2] Rachel M. Freathy,[3,4] Jeffrey C. Barrett,[2] Beverley Shields,[4] Andrew P. Morris,[2] Sian Ellard,[4,6] Christopher J. Groves,[1] Lorna W. Harries,[4] Jonathan L. Marchini,[7] Katharine R. Owen,[1] Beatrice Knight,[4]

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]

Sciencexpress Report

A Common Variant in the *FTO* Gene Is Associated with Body Mass Index and Predisposes to Childhood and Adult Obesity

Timothy M. Frayling,[1,2*] Nicholas J. Timpson,[3,4*] Michael N. Weedon,[1,2*] Eleftheria Zeggini,[3,5*] Rachel M. Freathy,[1,2] Cecilia M. Lindgren,[3,5] John R. B. Perry,[1,2] Katherine S. Elliott,[3] Hana Lango,[1,2] Nigel W. Rayner,[3,5] Beverley Shields,[2] Lorna W. Harries,[2] Jeffrey C. Barrett,[3] Sian Ellard,[2,6] Christopher J. Groves,[5] Bridget Knight,[2] Ann-Marie Patch,[2,6] Andrew R. Ness,[7] Shah Ebrahim,[8] Debbie A. Lawlor,[9] Susan M. Ring,[9] Yoav Ben-Shlomo,[9] Marjo-Riitta Jarvelin,[10,11] Ulla Sovio,[10,11] Amanda J. Bennett,[5] David Melzer,[1,12] Luigi Ferrucci,[13] Ruth J. F. Loos,[14] Inês Barroso,[15] Nicholas J. Wareham,[14] Fredrik Karpe,[5] Katharine R. Owen,[5] Lon R. Cardon,[3] Mark Walker,[16] Graham A. Hitman,[17] Colin N. A. Palmer,[18] Alex S. F. Doney,[19] Andrew D. Morris,[19] George Davey-Smith,[4] The Wellcome Trust Case Control Consortium,[20] Andrew T. Hattersley,[1,2†‡] Mark I. McCarthy[3,5†]

12 April 2007

Top 10 Results From Combined Analysis

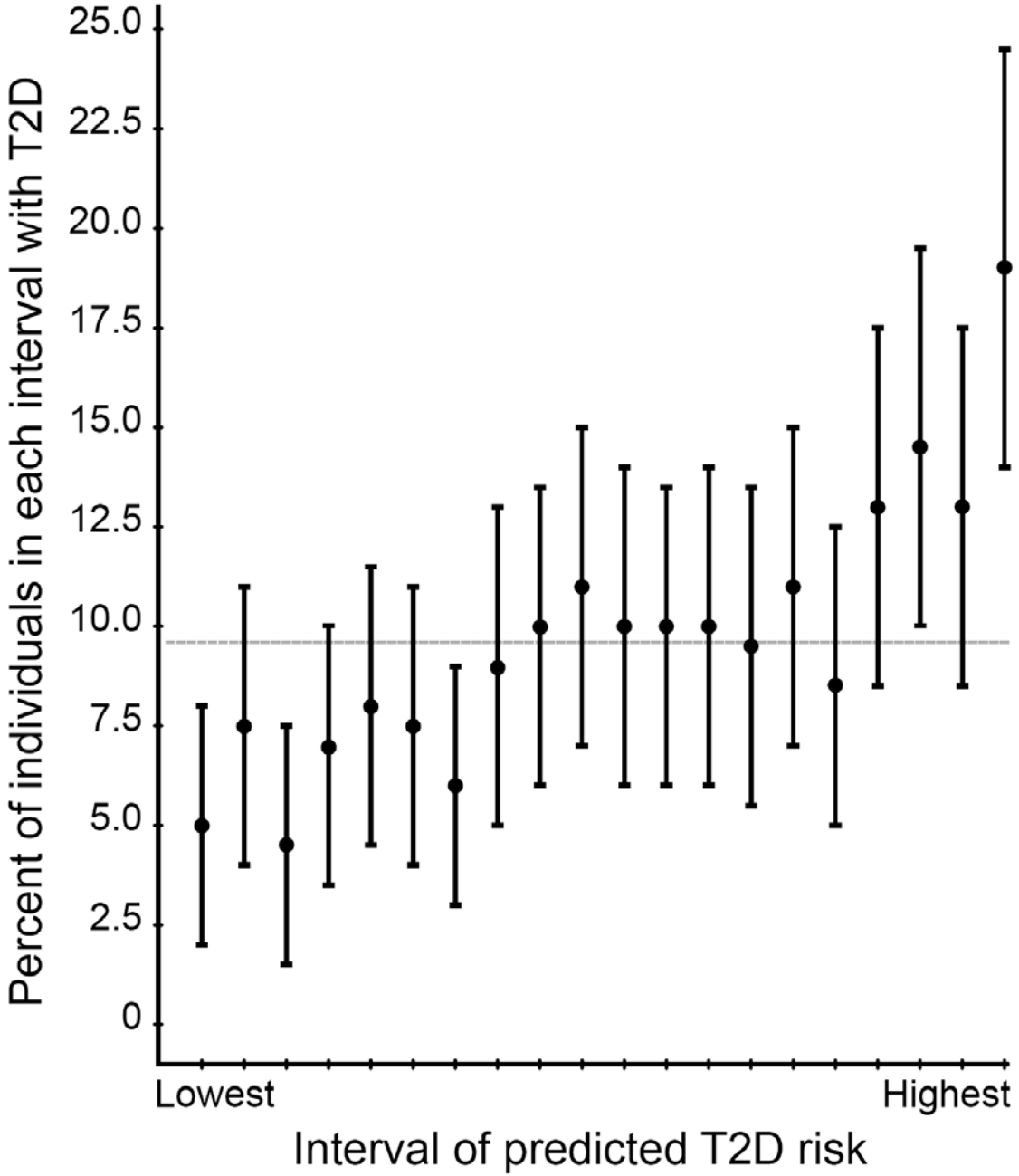
	FUSION		DGI		WTCCC/UKT2D		All Samples	
Gene	OR	p-value	OR	p-value	OR	p-value	OR	p-value
TCF7L2	1.34	1.3×10^{-8}	1.38	2.3×10^{-31}	1.37	6.7×10^{-13}	1.37	1.0×10^{-48}
IGF2BP2	1.18	2.1×10^{-4}	1.17	1.7×10^{-9}	1.11	1.6×10^{-4}	1.14	8.9×10^{-16}
CDKN2A/B	1.20	.0022	1.20	5.4×10^{-8}	1.19	4.9×10^{-7}	1.20	7.8×10^{-15}
FTO	1.11	0.016	1.03	0.25	1.23	7.3×10^{-14}	1.17	1.3×10^{-12}
CDKAL1	1.12	0.0095	1.08	0.0024	1.16	1.3×10^{-8}	1.12	4.1×10^{-11}
KCNJ11	1.11	0.013	1.15	1.0×10^{-7}	1.15	0.0013	1.14	6.7×10^{-11}
HHEX	1.10	0.026	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	0.047	1.12	7.0×10^{-5}	1.12	5.3×10^{-8}
Chr 11	1.48	5.7×10^{-8}	1.16	0.12	1.13	0.068	1.23	4.3×10^{-7}
PPARG	1.20	0.0014	1.09	0.019	1.23	0.0013	1.14	1.7×10^{-6}

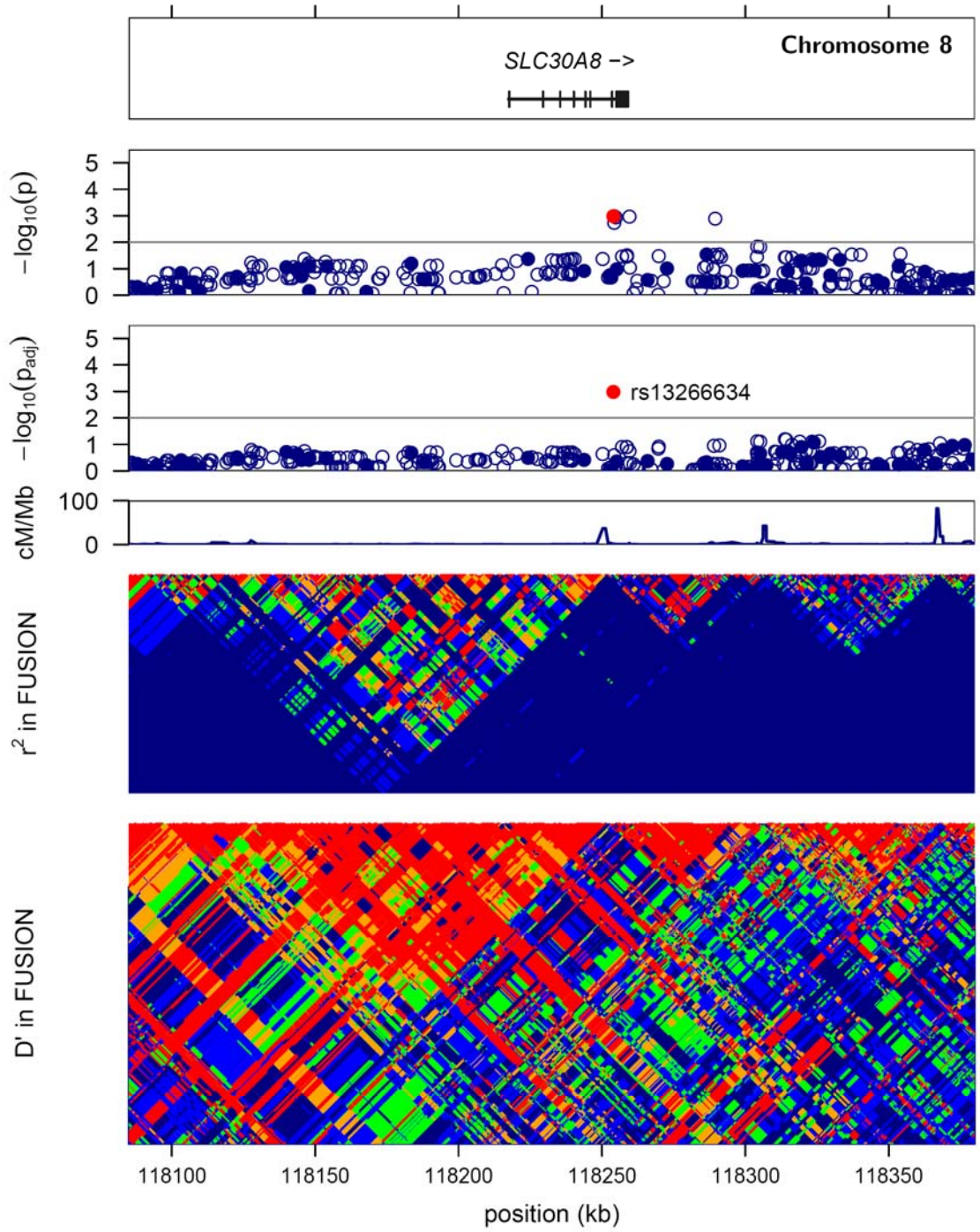
A variant in *CDKAL1* influences insulin response and risk of type 2 diabetes

Valgerdur Steinthorsdottir[1,15], Gudmar Thorleifsson[1,15], Inga Reynisdottir[1], Rafn Benediktsson[2,3], Thorbjorg Jonsdottir[1], G Bragi Walters[1], Unnur Styrkarsdottir[1], Solveig Gretarsdottir[1], Valur Emilsson[1], Shyamali Ghosh[1], Adam Baker[1], Steinunn Snorradottir[1], Hjordis Bjarnason[1], Maggie C Y Ng[4], Torben Hansen[5], Yu Bagger[6], Robert L Wilensky[7], Muredach P Reilly[7], Adebowale Adeyemo[8], Yuanxiu Chen[8], Jie Zhou[8], Vilmundur Gudnason[3], Guanjie Chen[8], Hanxia Huang[8], Kerrie Lashley[8], Ayo Doumatey[8], Wing-Yee So[4], Ronald C Y Ma[4], Gitte Andersen[5], Knut Borch-Johnsen[5,9,10], Torben Jorgensen[10], Jana V van Vliet-Ostaptchouk[11], Marten H Hofker[11,12], Cisca Wijmenga[13,14], Claus Christiansen[6], Daniel J Rader[7], Charles Rotimi[8], Mark Gurney[1], Juliana C N Chan[4], Oluf Pedersen[5,9],Gunnar Sigurdsson[2,3], Jeffrey R Gulcher[1], Unnur Thorsteinsdottir[1], Augustine Kong[1] & Kari Stefansson[1]

Top 10 Results From Combined Analysis

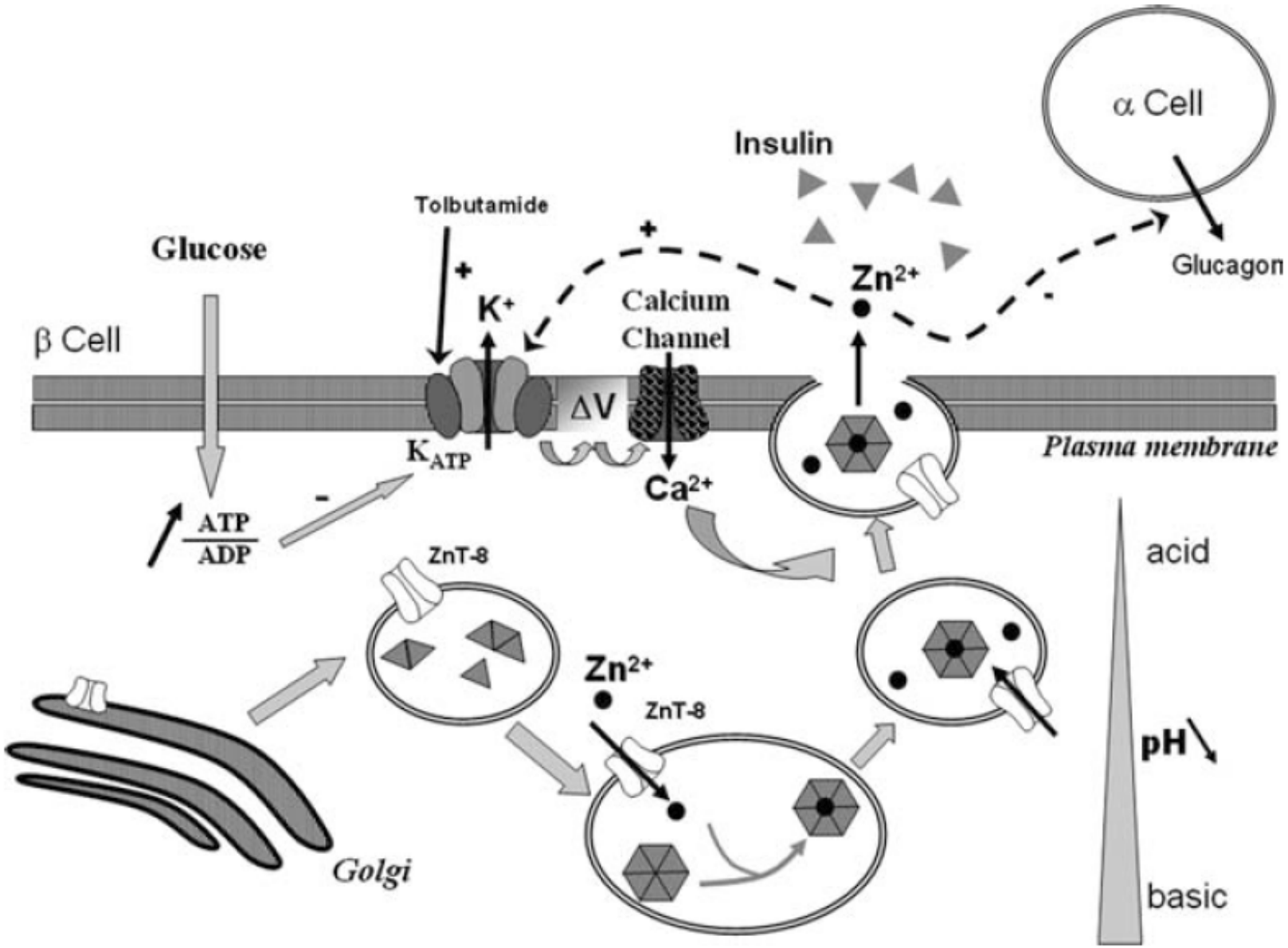
	FUSION		DGI		WTCCC/UKT2D		All Samples	
Gene	OR	p-value	OR	p-value	OR	p-value	OR	p-value
TCF7L2	1.34	1.3×10^{-8}	1.38	2.3×10^{-31}	1.37	6.7×10^{-13}	1.37	1.0×10^{-48}
IGF2BP2	1.18	2.1×10^{-4}	1.17	1.7×10^{-9}	1.11	1.6×10^{-4}	1.14	8.9×10^{-16}
CDKN2A/B	1.20	.0022	1.20	5.4×10^{-8}	1.19	4.9×10^{-7}	1.20	7.8×10^{-15}
FTO	1.11	0.016	1.03	0.25	1.23	7.3×10^{-14}	1.17	1.3×10^{-12}
CDKAL1	1.12	0.0095	1.08	0.0024	1.16	1.3×10^{-8}	1.12	4.1×10^{-11}
KCNJ11	1.11	0.013	1.15	1.0×10^{-7}	1.15	0.0013	1.14	6.7×10^{-11}
HHEX	1.10	0.026	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	0.047	1.12	7.0×10^{-5}	1.12	5.3×10^{-8}
Chr 11	1.48	5.7×10^{-8}	1.16	0.12	1.13	0.068	1.23	4.3×10^{-7}
PPARG	1.20	0.0014	1.09	0.019	1.23	0.0013	1.14	1.7×10^{-6}


25.0
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0
Percent of individuals in each interval with T2D
Lowest
Highest
Interval of predicted T2D risk

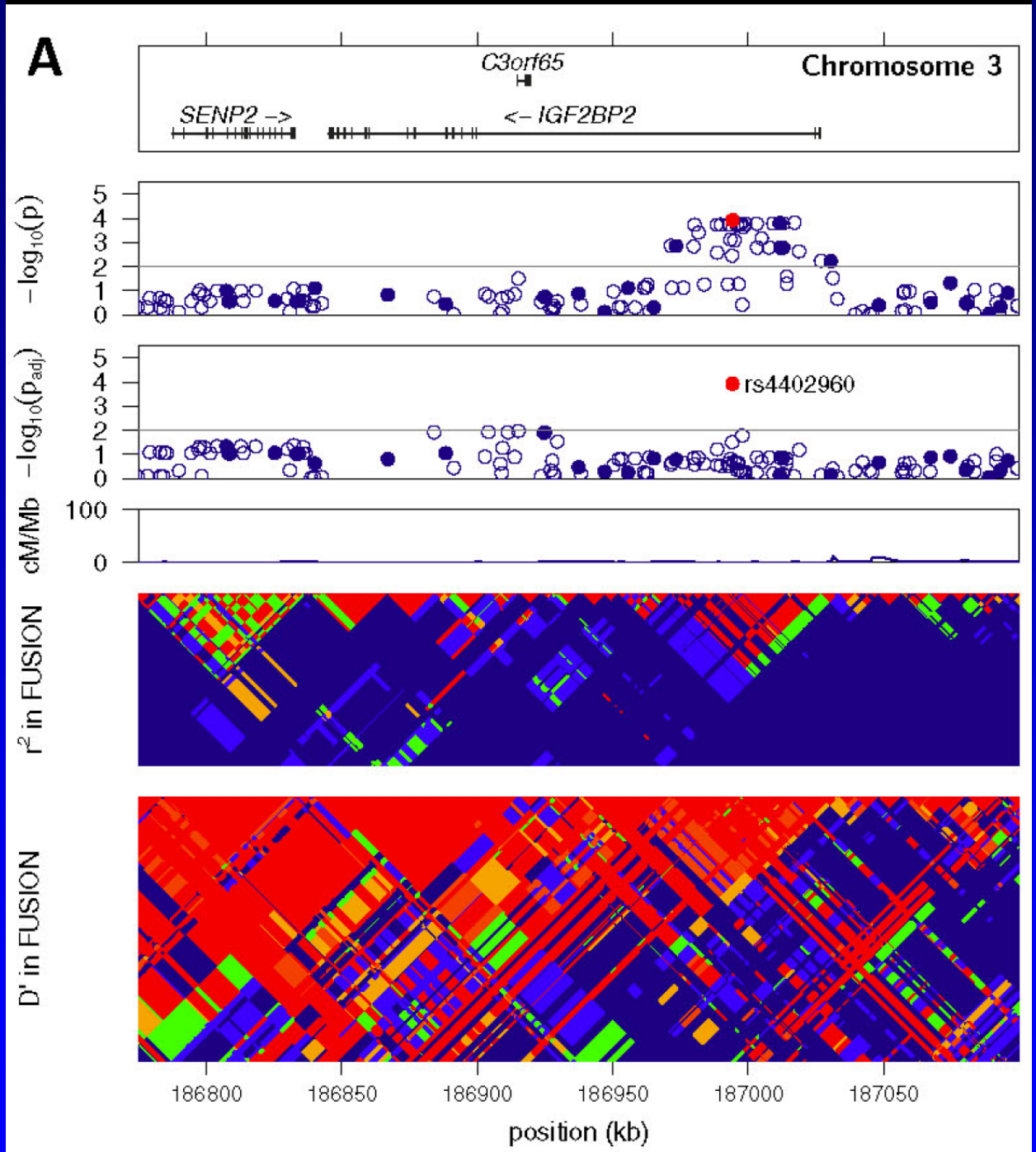


rs13266634 is in the coding region of SLC30A8, and changes a highly conserved arginine to a tryptophan

SLC30A8 - Beta Cell Zinc Transporter

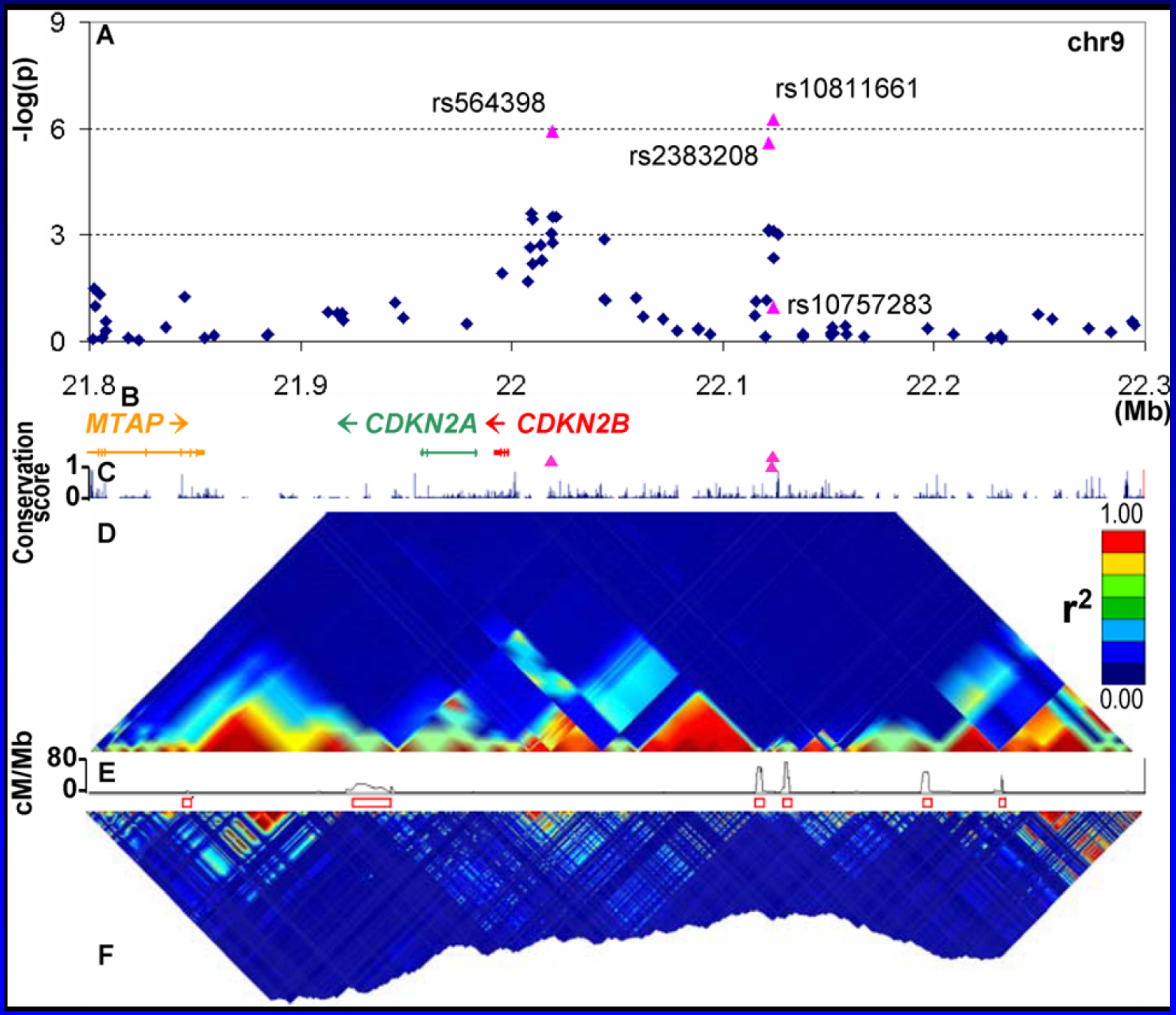


Chimienti (2005) BioMetals 18:313


A
C3orf65
Chromosome 3
SENP2 ->
<- IGF2BP2
−log10(p)
−log10(padj)
rs4402960
cM/Mb
r2 in FUSION
D' in FUSION
186800
186850
186900
186950
187000
187050
position (kb)

How could IGF2BP2 be related to diabetes?

- ***IGF2BP2* codes for the insulin-like growth factor 2 mRNA binding protein 2**
- **But we know very little about what this does**
- **A related gene, *IGF2BP1*, codes for a protein that binds to the upstream leader sequence of the insulin-like growth factor 2 (IGF2) mRNA and regulates IGF2 protein production**
- **IGF2 is involved in development, growth, and stimulation of insulin action**


A
chr9
-log(p)
9
6
3
0
rs564398
rs10811661
rs2383208
rs10757283
21.8
21.9
22
22.1
22.2
22.3
(Mb)
B
MTAP
CDKN2A
CDKN2B
Conservation score
1
0
C
D
1.00
r2
0.00
cM/Mb
80
0
E
F

Sciencexpress Report

A Common Allele on Chromosome 9 Associated with Coronary Heart Disease

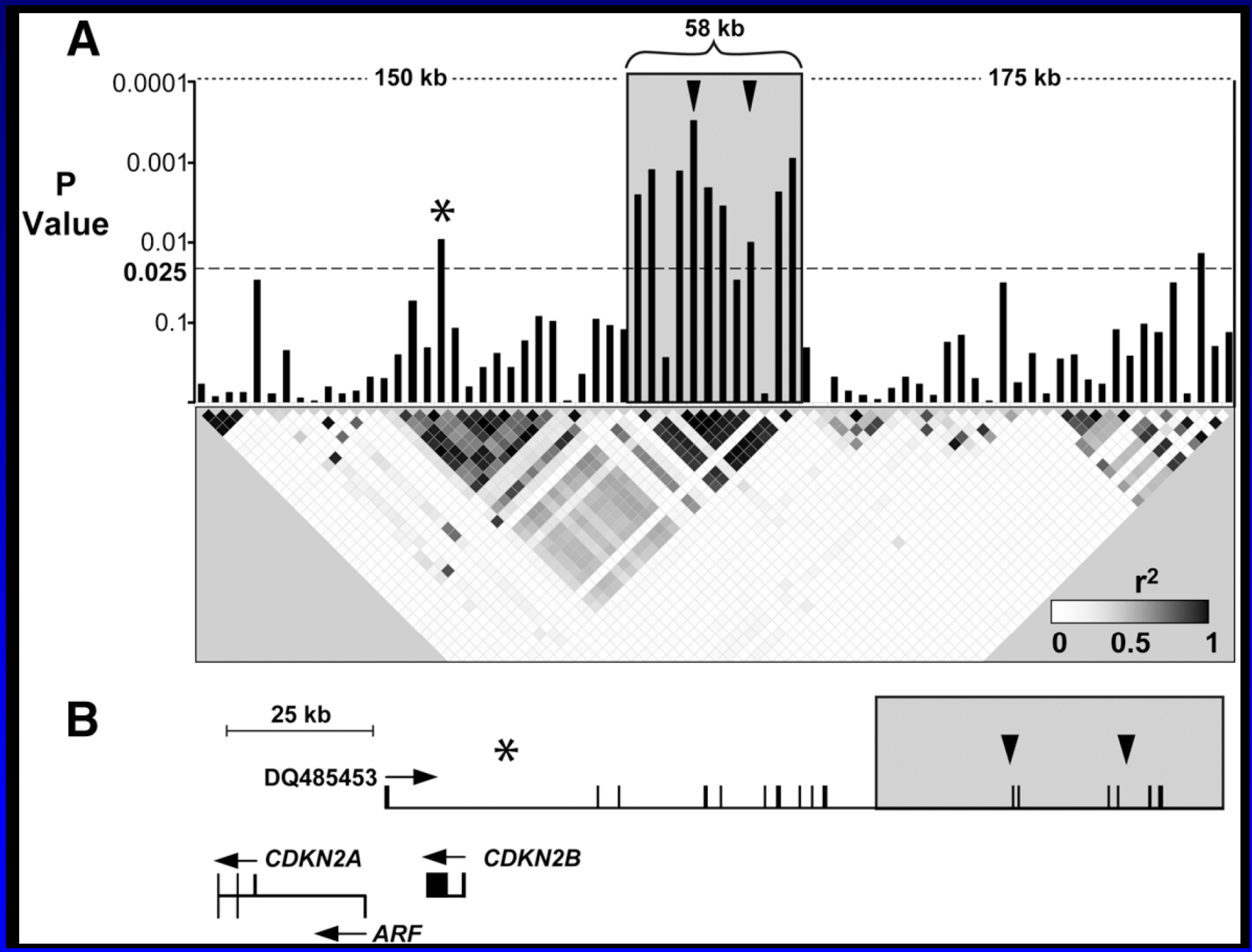
Ruth McPherson,[1]*† Alexander Pertsemlidis,[2]* Nihan Kavaslar,[1] Alexandre Stewart,[1] Robert Roberts,[1] David R. Cox,[3] David A. Hinds,[3] Len A. Pennacchio,[4] Anne Tybjaerg-Hansen,[5] Aaron R. Folsom,[6] Eric Boerwinkle,[7] Helen H. Hobbs,[2,9] Jonathan C. Cohen[2,8]†

[1]Division of Cardiology, University of Ottawa Heart Institute, Ottawa K1Y4W7, Canada. [2]Donald W. Reynolds Cardiovascular Clinical Research Center and the Eugene McDermott Center for Human Growth and Development, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA. [3]Perlegen Sciences, Mountain View, CA 94043; USA. [4]Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA & U.S. Department of Energy Joint Genome Institute, Walnut Creek, CA 94598, USA. [5]Department of Clinical Biochemistry, Rigshospitalet, Copenhagen University Hospital, Copenhagen DK-2100, Denmark. [6]Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN 55454, USA. [7]Human Genetics Center and Institute for Molecular Medicine, University of Texas Health

Sciencexpress Report

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 Gudbjartsson,[1] Kristinn P. Magnusson,[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] Arnaldur Gylfason,[1] Viola Vaccarino,[3] W. Craig Hooper,[3] Muredach P. Reilly,[4] Christopher B. Granger,[5] Harland Austin,[3] Daniel J Rader,[4] Svati H. Shah,[5] Arshed A. Quyyumi,[3] Jeffrey R. Gulcher,[1] Gudmundur Thorgeirsson,[3] Unnur Thorsteinsdottir,[1] Augustine Kong,[1]† Kari Stefansson[1]†


A
58 kb
150 kb
175 kb
0.0001
0.001
0.01
0.025
0.1
P
Value
r2
0
0.5
1
B
25 kb
DQ485453
CDKN2A
CDKN2B
ARF

Conclusions

- Early sharing of data with other groups was essential to successful discovery of risk variants
- Having “positive controls” was very helpful
- Merging genotype data across different GWA platforms is essentially a solved problem
- Delaying publication until the evidence was overwhelming was a wise choice
- These same data sets will no doubt reveal further gene variants that contribute to:
  - Type 2 diabetes
  - Related traits – lipids, BP, anthropometrics,..
- Principles for GWA publication will soon appear

What Constitutes Replication of a Genotype-Phenotype Association? Summary of an NCI-NHGRI Working Group

NCI-NHGRI Working Group on Replication in Association Studies

[*Nature*, in press]