

# UC San Diego

## UC San Diego Previously Published Works

### Title

Comprehensive Association Study of Type 2 Diabetes and Related Quantitative Traits With 222 Candidate Genes

### Permalink

<https://escholarship.org/uc/item/4dn2r5h4>

### Journal

Diabetes, 57(11)

### ISSN

0012-1797

### Authors

Gaulton, Kyle J  
Willer, Cristen J  
Li, Yun  
et al.

### Publication Date

2008-11-01

### DOI

10.2337/db07-1731

Peer reviewed

# Comprehensive Association Study of Type 2 Diabetes and Related Quantitative Traits With 222 Candidate Genes

Kyle J. Gaulton,<sup>1</sup> Cristen J. Willer,<sup>2</sup> Yun Li,<sup>2</sup> Laura J. Scott,<sup>2</sup> Karen N. Conneely,<sup>2</sup> Anne U. Jackson,<sup>2</sup> William L. Duren,<sup>2</sup> Peter S. Chines,<sup>3</sup> Narisu Narisu,<sup>3</sup> Lori L. Bonnycastle,<sup>3</sup> Jingchun Luo,<sup>4</sup> Maurine Tong,<sup>3</sup> Andrew G. Sprau,<sup>3</sup> Elizabeth W. Pugh,<sup>5</sup> Kimberly F. Doheny,<sup>5</sup> Timo T. Valle,<sup>6</sup> Gonçalo R. Abecasis,<sup>2</sup> Jaakko Tuomilehto,<sup>6,7,8</sup> Richard N. Bergman,<sup>9</sup> Francis S. Collins,<sup>3</sup> Michael Boehnke,<sup>2</sup> and Karen L. Mohlke<sup>1</sup>

**OBJECTIVE**—Type 2 diabetes is a common complex disorder with environmental and genetic components. We used a candidate gene–based approach to identify single nucleotide polymorphism (SNP) variants in 222 candidate genes that influence susceptibility to type 2 diabetes.

**RESEARCH DESIGN AND METHODS**—In a case-control study of 1,161 type 2 diabetic subjects and 1,174 control Finns who are normal glucose tolerant, we genotyped 3,531 tagSNPs and annotation-based SNPs and imputed an additional 7,498 SNPs, providing 99.9% coverage of common HapMap variants in the 222 candidate genes. Selected SNPs were genotyped in an additional 1,211 type 2 diabetic case subjects and 1,259 control subjects who are normal glucose tolerant, also from Finland.

**RESULTS**—Using SNP- and gene-based analysis methods, we replicated previously reported SNP-type 2 diabetes associations in *PPARG*, *KCNJ11*, and *SLC2A2*; identified significant SNPs in genes with previously reported associations (*ENPP1* [rs2021966,  $P = 0.00026$ ] and *NRF1* [rs1882095,  $P = 0.00096$ ]); and implicated novel genes, including *RAPGEF1* (rs4740283,  $P = 0.00013$ ) and *TP53* (rs1042522, Arg72Pro,  $P = 0.00086$ ), in type 2 diabetes susceptibility.

**CONCLUSIONS**—Our study provides an effective gene-based approach to association study design and analysis. One or more of the newly implicated genes may contribute to type 2 diabetes pathogenesis. Analysis of additional samples will be necessary to determine their effect on susceptibility. *Diabetes* 57:3136–3144, 2008

From the <sup>1</sup>Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina; the <sup>2</sup>Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan; the <sup>3</sup>Genome Technology Branch, National Human Genome Research Institute, Bethesda, Maryland; the <sup>4</sup>Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina; the <sup>5</sup>Center for Inherited Disease Research, Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, Maryland; the <sup>6</sup>Diabetes and Genetic Epidemiology Unit, Department of Epidemiology and Health Promotion, National Public Health Institute, Helsinki, Finland; the <sup>7</sup>Department of Public Health, University of Helsinki, Helsinki, Finland; the <sup>8</sup>South Ostrobothnia Central Hospital, Seinäjoki, Finland; and the <sup>9</sup>Department of Physiology and Biophysics, Keck School of Medicine, University of Southern California, Los Angeles, California.

Corresponding author: Karen Mohlke, mohlke@med.unc.edu.

Received 11 December 2007 and accepted 21 July 2008.

Published ahead of print at <http://diabetes.diabetesjournals.org> on 4 August 2008. DOI: 10.2337/db07-1731.

© 2008 by the American Diabetes Association. Readers may use this article as long as the work is properly cited, the use is educational and not for profit, and the work is not altered. See <http://creativecommons.org/licenses/by-nc-nd/3.0/> for details.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

**T**ype 2 diabetes is a metabolic disorder characterized by insulin resistance and pancreatic  $\beta$ -cell dysfunction and is a leading cause of morbidity and mortality in the U.S. and worldwide. The incidence of type 2 diabetes is rapidly increasing, with 1.6 million new cases of diabetes diagnosed in individuals aged  $\geq 20$  years in the U.S. in 2007 (available at <http://www.diabetes.niddk.nih.gov/dm/pubs/statistics/>). While environmental factors play a major role in predisposition to type 2 diabetes, substantial evidence supports the influence of genetic factors on disease susceptibility. For example, the twin concordance rate is an estimated 34% for monozygotic twins and 16% for dizygotic twins (1). However, the underlying genetic variants are just beginning to be identified (2).

Numerous published reports (3–5) have identified association between type 2 diabetes and common genetic variants in human populations; however, until very recently, variants in only a few genes have been consistently replicated across populations and with large sample sizes. Among these are the Pro12Ala (rs1801282) variant in peroxisome proliferator–activated receptor  $\gamma$  (*PPARG*) (6), the Glu23Lys (rs5210) variant in the potassium channel gene *KCNJ11* (7), and several variants in the Wnt-receptor signaling pathway member *TCF7L2* (8).

Recent genome-wide studies have implicated many previously unreported genes in type 2 diabetes susceptibility. The first reported genome-wide association (GWA) scan implicated variants at five susceptibility loci that include *TCF7L2* and novel loci near the genes *SLC30A8*, *IDE-KIF11-HHEX*, *LOC387761*, and *EXT-ALX4* (9). Three companion GWA studies (10–12), including one by our group, replicated evidence for *PPARG*, *KCNJ11*, *TCF7L2*, *SLC30A8*, and *IDE-KIF11-HHEX* and provided new evidence for *CDKAL1*, *CDKN2A-CDKN2B*, *IGF2BP2*, *FTO*, and a region of chromosome 11 with no annotated genes. Additional GWA studies (13–18) provided additional evidence for *TCF7L2*, *CDKAL1*, and *SLC30A8*. The candidate genes *WFS1* (19) and *TCF2* (20,21) have also been confirmed in large samples, bringing the current list of type 2 diabetes susceptibility loci to at least 10. The recent discovery of these loci still explains only a small fraction (~2.3%) of the overall risk of type 2 diabetes (12). Therefore, novel susceptibility genes remain to be identified through increasingly comprehensive analyses of both individual genes and the entire genome.

The Finland-U.S. Investigation of Type 2 Diabetes Genetics (FUSION) study aims to identify variants influenc-

TABLE 1  
Characteristics of the stage 1 and 2 case and control samples

	Stage 1		Stage 2	
	Case subjects	Control subjects	Case subjects	Control subjects
<i>n</i>	1,161	1,174	1,215	1,258
Male subjects	653	574	724	768
Female subjects	508	600	491	490
Age of diagnosis (years)	53.0 ± 12.0	N/A	56.0 ± 12.0	N/A
Study age (years)	63.4 ± 11.2	64 ± 11.7	60.0 ± 11.5	59.0 ± 10.6
BMI (kg/m <sup>2</sup> )	29.8 ± 6.1	26.8 ± 5.0	30.1 ± 6.7	26.4 ± 4.9
Fasting glucose (mmol/l)	8.4 ± 3.9	5.4 ± 0.7	7.2 ± 2.1*	5.4 ± 0.6†

Data are medians ± interquartile ranges unless otherwise indicated. \**n* = 204 and †*n* = 583 values converted from whole blood to plasma glucose equivalent using a prediction equation from the European Diabetes Epidemiology Group, of which †*n* = 262 fasted <8 h.

ing susceptibility to type 2 diabetes and related quantitative traits in the Finnish population (22). FUSION has previously identified modest type 2 diabetes association in Finns with variants in *HNF4A* (23); four genes known to cause maturity-onset diabetes of the young (5,23,24); *PPARG*, *KCNJ11*, *ENPP1*, *SLC2A2*, *PCK1*, *TNF*, *IL6* (5), and *TCF7L2* (25); and the loci identified in the GWA studies.

As a complementary approach to GWA studies, which are conducted without a priori biological hypotheses, we sought to perform an in-depth analysis of >200 genes likely to influence susceptibility to type 2 diabetes and quantitative trait variation that we selected by applying CandidAtE Search And Rank (CAESAR), a text- and data-mining algorithm (26). We aimed to analyze the full spectrum of HapMap-based common variation in each of these candidate genes. The combination of high throughput genotyping, linkage disequilibrium (LD) information from HapMap (27), the ability to impute ungenotyped variants (28), and the improved functional annotation of the genome makes in-depth candidate gene-based association analysis possible.

## RESEARCH DESIGN AND METHODS

The stage 1 sample set consisted of 2,335 Finnish individuals from the FUSION (22,29) and Finrisk 2002 (30) studies (Table 1) (online appendix Table 1A [available at <http://dx.doi.org/10.2337/db07-1731>]). The sample included 1,161 individuals with type 2 diabetes and 1,174 control subjects with normal glucose tolerance. Diabetes was defined according to 1999 World Health Organization criteria (fasting plasma glucose concentration ≥7.0 mmol/l or 2-h plasma glucose concentration ≥11.1 mmol/l), by report of diabetes medication use, or based on medical record review. Normal glucose tolerance was defined as having fasting glucose <6.1 mmol/l and 2-h glucose <7.8 mmol/l. A total of 120 FUSION offspring with genotyped parents were included for quantitative trait analysis; all offspring had normal glucose tolerance except one type 2 diabetic individual who was included in the case sample.

Stage 2 consisted of 2,473 Finnish individuals (Table 1) (online appendix Table 1B) and included 1,215 individuals with type 2 diabetes and 1,258 control subjects with normal glucose tolerance (10). A total of 56 duplicate samples were used for quality control. The sample sets are identical to those used in the FUSION GWA study (10). Study protocols were approved by local ethics committees and/or institutional review boards, and informed consent was obtained from all study participants.

**Gene selection.** A total of 222 candidate genes were selected for study using two strategies. Two hundred and seventeen candidate genes were selected using CAESAR, an algorithm that prioritizes candidate genes for complex human traits based on trait-relevant functional annotation (26). Given a trait-relevant input text, CAESAR 1) uses text mining to extract gene symbols and to find and rank terms present in four biomedical ontologies (gene ontology biological process [31], gene ontology molecular function [31], eVOC anatomy [32], and mammalian phenotype ontology [33]) based on frequency of occurrence, 2) uses the ranked ontology terms and extracted gene symbols to

data mine several public databases for human genes annotated with the ontology terms or extracted gene symbols, and 3) integrates the resulting gene annotation lists to provide a combined score and rank for each gene. Details of gene selection using custom parameters for CAESAR are provided in the online appendix.

Five genes were not ranked high enough to have been included using CAESAR. *ENPP1*, *HFE*, *WFS1*, and *ZNHIT3* were included because each had one or more single nucleotide polymorphisms (SNPs) associated with type 2 diabetes ( $P < 0.1$ ) in a prior study of a subset of FUSION samples (6) (C.J.W., L.L.B., M.B., and K.L.M., unpublished data); in addition, *ENPP1* and *WFS1* had been previously studied as type 2 diabetes candidate genes. *CAPN10* was included because it had been previously studied by FUSION (34) and others (35,36).

**SNP selection.** We defined the “transcribed region” of each of the 222 candidate genes as the sequence including the first exon of any transcribed isoform through the last exon of any transcribed isoform, and we aimed to capture variation up to 10 kb upstream and 5 kb downstream of the transcribed region (−10 kb/+5 kb). In this process, we allowed SNPs to be located as far as 50 kb upstream and 50 kb downstream (−50 kb/+50 kb) of the transcribed region if they tagged a −10-kb/+5-kb SNP at  $r^2 > 0.8$ .

Briefly, 3,531 SNPs were selected for stage 1 genotyping as follows. We selected SNPs from the Illumina Infinium II HumanHap300 BeadChip that tagged one or more −10-kb/+5-kb SNPs ( $r^2 > 0.8$ ). Then, to evaluate each gene region more comprehensively, we selected 1) additional tagSNPs and 2) functionally annotated non-HapMap SNPs for genotyping on an Illumina GoldenGate panel. We also included eight SNPs that had been previously genotyped in candidate gene studies on a smaller subset of FUSION samples (5). Additional details of SNP selection are provided in the online appendix.

**Genotyping.** Stage 1 genotyping of 317,503 SNPs was performed at the Center for Inherited Disease Research on the HumanHap300 BeadChip using the Illumina Infinium II assay protocol (10), and 1,527 SNPs were genotyped in partnership with the Mammalian Genotyping Core at the University of North Carolina using the Illumina GoldenGate assay. We performed additional genotyping for eight previously reported SNPs (5) using the Sequenom homogeneous MassEXTEND assay and four imputed SNPs using Applied Biosystems TaqMan allelic discrimination assays. There was a genotype consistency rate of >99.88% between each platform, using 79 duplicate samples. Stage 2 genotyping of 31 SNPs was performed using the homogeneous MassEXTEND assay; there was a genotype consistency rate of 100%, using 56 duplicate samples. SNP and sample success rates and quality-control filters are described in the online appendix.

**Imputation.** We used MACH, a computationally efficient hidden Markov model-based algorithm (available at <http://www.sph.umich.edu/csg/abecasis/MACH/>) (28), to impute genotypes in FUSION samples for 7,498 common (minor allele frequency [MAF] > 0.05) HapMap SNPs present in the target regions but not genotyped in our study. To improve the quality of imputation near the ends of the target regions, we used at least 1 Mb of flanking genotype information to impute SNPs in target regions.

**Coverage of HapMap SNPs.** Coverage was calculated as the percentage of all common (MAF > 0.05) HapMap Release 21 CEU SNPs in the −10-kb/+5-kb gene regions that are tagged by a genotyped SNP at an  $r^2$  threshold of at least 0.8.

**Type 2 diabetes association analysis.** Genotyped SNPs were tested for type 2 diabetes association using logistic regression under additive ( $P_{\text{add}}$ ), dominant, and recessive genetic models with adjustment for 5-year age category, sex, and birth province. Imputed SNPs were tested for type 2 diabetes association using logistic regression under an additive model ( $P_{\text{impute}}$ ), with

TABLE 2  
Coverage of 10,762 HapMap SNPs (MAF > 0.05)\* within -10 kb/+5 kb of 222 candidate genes

	SNPs -10 kb/+5 kb of gene		
	Number of SNPs analyzed†	Number captured‡	Percent captured‡
SNPs genotyped on GWA panel only	2,150	8,507	79.04
All 3,531 genotyped SNPs	3,531	10,299	95.74
Genotyped and imputed SNPs from GWA panel only	10,596	10,647	98.93
All 3,531 genotyped and 7,498 imputed SNPs	11,029	10,752§	99.91

\*MAF > 0.05 in HapMap CEU. †Genotyped SNPs are located within -50 kb/+50 kb of a gene but may not be within -10 kb/+5 kb of a gene. Imputed SNPs are all located within -10 kb/+5 kb of a gene. ‡HapMap SNPs genotyped, imputed, or tagged ( $r^2 > 0.8$ ) by a genotyped SNP. §A total of 10,752 includes 3,187 genotyped SNPs, 7,498 imputed SNPs, and 67 SNPs tagged ( $r^2 > 0.8$ ) by a genotyped SNP.

the expected allele count in place of the allele count and adjusted for the same covariates. This approach takes into account the degree of uncertainty of genotype imputation in a computationally efficient manner by replacing allele counts (0, 1, and 2) at the marker locus by predicted allele counts based on estimated probabilities of 0, 1, or 2 copies of a SNP allele (available at <http://www.sph.umich.edu/csg/abecasis/MACH/>) (28).

We accounted for carrying out multiple correlated tests using the  $P$  value adjusted for correlated tests ( $P_{ACT}$ ) method (37). The  $P_{ACT}$  method was used to correct the minimum  $P$  value among 1) tests of three genetic models for a single SNP ( $P_{SNP}$ ) and 2) multiple SNPs and models across a gene region ( $P_{gene}$ ). Details are provided in the online appendix. We determined the independence of significant association signals in genes by including one SNP as a covariate in logistic regression and reassessing the evidence for association with the other SNPs.

**Quantitative trait analysis.** We tested all genotyped and imputed SNPs for association with 20 type 2 diabetes-related quantitative traits, including, in control subjects only, fasting insulin, fasting glucose, homeostasis model adjustment, and fasting free fatty acids; and, in all samples, BMI, weight, waist circumference, hip circumference, waist-to-hip ratio, waist-to-height<sup>2</sup> ratio, total cholesterol, HDL cholesterol, LDL cholesterol, triglyceride level, cholesterol-to-HDL ratio, triglyceride-to-HDL ratio, diastolic blood pressure, systolic blood pressure, pulse, and pulse pressure.

For case and control subjects separately, we regressed the quantitative trait variables on age, age<sup>2</sup>, sex, birth province, and study indicator and transformed the residuals of each quantitative trait to approximate normality using inverse normal scores, which involves ranking the residual values and then converting these to z-scores according to quantiles of the standard normal distribution. We then carried out association analysis on the residuals. To allow for relatedness, regression coefficients were estimated in the context of a variance component model that also accounted for background polygenic effects (38). For genotyped SNPs, we tested for association using the residuals under an additive model. For imputed SNPs, we tested for association using the residuals and the expected allele count in place of the allele count under an additive model. Case and control results were combined using meta-analysis, as described in the online appendix.

## RESULTS

We studied 222 candidate genes for type 2 diabetes association in our stage 1 sample of 1,161 type 2 diabetic case subjects and 1,174 control subjects with normal glucose tolerance from the FUSION study (Table 1). Of 10,762 target HapMap SNPs (MAF > 0.05) in the -10-kb/+5-kb gene regions, 3,531 genotyped SNPs cover 10,299 (95.7%) SNPs at an  $r^2$  threshold of 0.8. This represents an improvement over the genome-wide HumanHap300 genotyped SNPs, which alone cover 79.0% of the target SNPs at  $r^2 \geq 0.8$  (Table 2). A total of 3,187 of 3,531 genotyped SNPs are

located in the -10-kb/+5-kb regions. Of the remaining 7,575 ungenotyped target SNPs, 7,498 were successfully imputed. Altogether, 99.9% of all target variation was genotyped, imputed, or tagged ( $r^2 \geq 0.8$ ) by an analyzed SNP.

We evaluated the significance of genotyped SNPs in each gene region after correcting for multiple SNPs tested while accounting for the LD between SNPs, designated  $P_{gene}$  (37). Given six pairs of adjacent genes (see online appendix), we analyzed 216 distinct gene regions for type 2 diabetes association (online appendix Table 2). SNPs in four gene regions (rs11183212 in *ARID2* [ $P_{gene} = 0.0029$ ], rs2235718 in *FOXCI* [ $P_{gene} = 0.0028$ ], rs8069976 in *SOCS3* [ $P_{gene} = 0.0037$ ], and rs222852 in *SLC2A4* [ $P_{gene} = 0.0024$ ]) were significantly associated with type 2 diabetes at  $P_{gene} < 0.005$ , although no  $P_{gene}$  result reached a study-wide significance of 0.00023, a threshold determined using a Bonferroni correction. SNPs in 19 genes were significant at  $P_{gene} < 0.05$ , including SNPs in three genes previously implicated in type 2 diabetes susceptibility in FUSION (5) (Table 3). There was an excess of significant  $P_{gene}$  results at both thresholds (4 at  $P_{gene} < 0.005$  [ $P = 0.024$ ]; 19 at  $P_{gene} < 0.05$  [ $P = 0.013$ ]). The excess of significant results at  $P_{gene} < 0.005$  is maintained after excluding 1) seven genes showing prior evidence of association with any SNP in FUSION samples ( $P = 0.022$ ) or 2) five genes not selected by CAESAR ( $P = 0.022$ ), as no excluded genes were significant at that threshold (see online appendix).

To evaluate all 3,531 genotyped SNPs (online appendix Table 3), we permuted the case/control status to estimate whether an excess of significant results was observed. A total of 214 SNPs showed significant type 2 diabetes association at a  $P_{SNP}$  threshold of 0.05, and, of these, 26 were associated at a  $P_{SNP}$  threshold of 0.005 (Table 4). There was modest, but not significant, excess at both of these  $P_{SNP}$  thresholds (observed = 214, expected = 183.3,  $P = 0.09$  and observed = 26, expected = 18.9,  $P = 0.12$ , respectively). The most significant  $P_{SNP}$  value of  $3.6 \times 10^{-4}$  was observed for rs11183212, an intronic SNP in the *ARID2* gene, but when compared with an empirical distribution of the most significant  $P$  values, this SNP does not reach a study-wide significance threshold of  $6.3 \times 10^{-5}$ , based on 1,000 permutations. In the combined stage 1 and 2 sample, we have >99% power (80% in stage 1 alone) to detect the most strongly associated previously observed type 2 diabetes SNP, rs7903146 in *TCF7L2* (9–12), at a study-wide significance level, and substantially less power to detect type 2 diabetes-associated SNPs with smaller effect sizes.

Nineteen of 216 gene regions have at least one SNP significantly associated with type 2 diabetes at  $P_{SNP} < 0.005$ ; among these, Pro12Ala (rs1801282) in *PPARG* ( $P_{SNP} = 0.0025$ ) was the only SNP that matched or was in high LD ( $r^2 \geq 0.8$ ) with a previously reported variant, given the available HapMap LD information. Imputation identified 421 additional SNPs in 59 genes significantly associated with type 2 diabetes ( $P_{impute} < 0.05$ ) (online appendix Table 4), including SNPs in 10 genes that did not contain a significant genotyped SNP ( $P_{SNP} > 0.05$ ). We genotyped four of these initially imputed SNPs that were both significantly associated with type 2 diabetes ( $P_{impute} < 0.05$ ) and for which the imputation-based  $P$  value was at least five times more significant than that for any nearby genotyped SNP; three of four SNPs had

TABLE 3  
Gene regions (−10 kb/+5 kb) associated with type 2 diabetes ( $P_{\text{gene}} < 0.05$ ) in stage 1 samples

Gene symbol	Chromosome	Start position* (bp)	End position* (bp)	Coverage (%)†	SNP‡	$P_{\text{gene}}$
<i>SLC2A4</i>	17	7,125,835	7,131,125	90.0	rs222852	0.0024
<i>FOXC1</i>	6	1,555,680	1,557,341	100.0	rs2235718	0.0028
<i>ARID2</i>	12	44,409,887	44,588,086	97.8	rs11183212	0.0029
<i>SOCS3</i>	17	73,864,459	73,867,753	100.0	rs8069976	0.0037
<i>FOXC2</i>	16	85,158,443	85,159,948	100.0	rs4843165	0.012
<i>ENPP1</i> §¶	6	132,170,853	132,254,043	94.8	rs9402346	0.014
<i>PRKAA2</i>	1	56,823,041	56,886,142	90.0	rs11206883	0.014
<i>JAK3</i>	19	17,797,961	17,819,800	85.7	rs11888	0.016
<i>CBLB</i>	3	106,859,799	107,070,577	98.8	rs17280845	0.017
<i>SLC2A2</i> §	3	172,196,839	172,227,470	100.0	rs10513684	0.023
<i>PRKAR2B</i>	7	106,279,129	106,396,206	97.0	rs2395836	0.027
<i>EDF1</i>	9	137,032,408	137,036,575	100.0	rs3739942	0.029
<i>PCK2</i>	14	23,633,323	23,643,177	100.0	rs2759407	0.034
<i>PRKAG3</i>	2	219,512,611	219,522,017	100.0	rs6436094	0.037
<i>MECR</i>	1	29,340,001	29,378,070	87.1	rs10915239	0.038
<i>RXRA</i>	9	134,519,422	134,558,376	85.7	rs3118526	0.040
<i>PPARGC1A</i>	4	23,469,914	23,567,969	94.4	rs2970871	0.041
<i>PPARG</i> §	3	12,304,359	12,450,840	99.1	rs1801282	0.042
<i>NR1I3</i>	1	158,012,528	158,021,028	100.0	rs2502807	0.049

\*Start and end positions of transcribed region (see research design and methods). Positions based on hg17. †Percentage of common (MAF > 0.05) SNPs within −10 kb/+5 kb of a gene and captured at  $r^2$  of at least 0.8. ‡SNP with minimum  $P$  value in given gene used to calculate  $P_{\text{gene}}$  value (see RESEARCH DESIGN AND METHODS). §Gene has previous evidence of association in FUSION. ¶Selected for study only based on previous evidence of association in FUSION.

highly concordant imputed and genotyped  $P$  values (online appendix Table 5).

We selected for follow-up genotyping in stage 2 samples 24 SNPs that were either significant at  $P_{\text{SNP}} < 0.005$  or, if a nonsynonymous variant, significant at  $P_{\text{SNP}} < 0.01$  (Table 1). The most significant SNPs in the combined stage 1 and 2 samples were rs4740283 in *RAPGEF1* ( $P_{\text{SNP}} = 0.00013$ ), rs2021966 in *ENPP1* ( $P_{\text{SNP}} = 0.00026$ ), Arg72Pro (rs1042522) in *TP53* ( $P_{\text{SNP}} = 0.00086$ ), and rs1882095 in *NRF1* ( $P_{\text{SNP}} = 0.00096$ ). In total, 16 SNPs were significant at  $P_{\text{SNP}} < 0.05$  in the combined stage 1 and 2 samples (Table 4).

To evaluate the effect of BMI, we included BMI as an additional covariate in an analysis of the additive model for all genotyped and imputed SNPs. Of 11 SNPs originally significant at  $P_{\text{add}} < 0.001$ , all  $P$  values were similar ( $P_{\text{add}} < 0.01$ ) after adjustment (online appendix Table 6A). Of 16 SNPs significant at  $P_{\text{add}} < 0.001$  after adjustment, two SNPs had notably less significant  $P$  values ( $P_{\text{add}} > 0.01$ ) before adjustment; both SNPs are located at the *TRIP10/C3* locus (online appendix Table 6B).

Four genotyped and 30 imputed SNPs were strongly associated ( $P < 0.0001$ ) with one or more of 20 quantitative traits after combining case and control subjects by meta-analysis (see RESEARCH DESIGN AND METHODS) (Table 5 and online appendix Table 7). Variants in *APOE* and *PPARA* showed strong evidence of association with serum lipid levels, confirming previous reports (39,40). Strong novel associations ( $P < 1 \times 10^{-5}$ ) were observed for rs4912407 in *PRKAA2* with triglyceride level ( $P = 3.68 \times 10^{-6}$ ), rs10517844 in *CPE* with HDL level ( $P = 2.07 \times 10^{-5}$ ), and rs4689388 in *WFS1* with LDL level ( $P = 5.30 \times 10^{-5}$ ). We followed-up genotyped SNPs significantly associated ( $P < 0.0001$ ) with one or more quantitative traits by genotyping the stage 2 samples. No SNP showed study-wide significance in the combined stage 1 and 2 samples (Table 5).

## DISCUSSION

In this study, we evaluated the evidence for type 2 diabetes association for SNPs in 222 candidate genes and provided a framework for thorough analysis of association of common variation to disease using gene-based functional annotation, HapMap LD information, and imputation of genotypes. This framework could be used in the context of a GWA study or an independent investigation of candidate genes. We replicated previous type 2 diabetes association with SNPs in *PPARG*, *KCNJ11*, and *SLC2A2*; identified significant SNPs in genes previously implicated in type 2 diabetes risk, *NRF1* and *ENPP1*; and identified additional genes that may influence susceptibility to type 2 diabetes and related quantitative traits, including *RAPGEF1* and *TP53*. While some of the genes may be significant by chance, one or more may represent true susceptibility genes. We expect that true susceptibility genes identified in our sample set will, in many cases, be shared in additional populations, as the FUSION GWA study identified many of the same risk alleles as other GWA studies of European populations (9–13).

To assess the role of 222 genes in susceptibility to type 2 diabetes, we attempted to assess complete coverage of common (MAF > 0.05) SNPs in the HapMap CEU database. The coverage of common HapMap CEU SNPs across all 222 candidate genes using genotyped SNPs was 95.7%, a 16.7% percent improvement over the coverage of 79.0% based on the Illumina HumanHap300 genome-wide panel (Table 2). HapMap provides excellent coverage of common variation in European samples; however, there are additional non-HapMap SNPs in these gene regions (27). Of 122 genotyped SNPs not in HapMap, 10 were not tagged at an  $r^2$  threshold of 0.8 by a HapMap SNP, indicating that some of the non-HapMap variation is better covered in our study than the GWA study panel.

Our SNP that is most strongly associated with type 2

TABLE 4  
Type 2 diabetes association for SNPs genotyped in FUSION stage 1 and 2 samples, sorted by combined stages 1 and 2  $P_{\text{SNP}}$

SNP	Gene symbol	Chromosome	Position (bp)*	Risk/nonrisk allele	Risk allele frequency	Stage 1 $P_{\text{SNP}}$	Stage 2 $P_{\text{SNP}}$	Model	Combined stage 1 and 2		
									$P$ value	Odds ratio (95% CI)	$P_{\text{SNP}}$
rs4740283	RAPGEF1	9	131,477,850	G/A	0.104	0.0042	0.030	REC	0.000052	3.12 (1.73–5.63)	0.00013
rs2021966†	ENPPI	6	132,192,132	A/G	0.608	0.00018	0.27	REC	0.00010	1.27 (1.13–1.43)	0.00026
rs1042522†‡	TP53	17	7,520,197	G/C	0.263	0.010	0.067	MUL	0.00037	1.18 (1.08–1.30)	0.00086
rs1882095	NRF1	7	128,991,595	T/C	0.381	0.0036	0.061	DOM	0.00043	1.24 (1.10–1.40)	0.00096
rs10513684	SLC2A2	3	172,206,912	C/T	0.918	0.0046	0.20	MUL	0.00010	1.28 (1.11–1.49)	0.00023
rs1801282	PPARG	3	12,368,125	C/G	0.836	0.0025	0.44	MUL	0.00014	1.20 (1.07–1.33)	0.00034
rs222852	SLC2A4	17	7,081,330	A/G	0.610	0.00048	0.18	MUL	0.00029	1.14 (1.04–1.23)	0.00070
rs4843165	FOXO2	16	85,162,542	C/T	0.706	0.0038	0.28§	MUL	0.00033	1.15 (1.05–1.25)	0.00078
rs5400‡	SLC2A2	3	172,215,002	G/A	0.871	0.0065	0.46	MUL	0.00045	1.19 (1.06–1.35)	0.010
rs858341	ENPPI	6	132,202,148	G/A	0.510	0.0039	0.70§	REC	0.00052	1.21 (1.06–1.39)	0.012
rs1349498	RAPGEF4	2	173,418,113	C/T	0.729	0.0015	0.68	DOM	0.00065	1.35 (1.09–1.67)	0.015
rs8069976	SOC3	17	73,861,445	C/A	0.849	0.0011	0.90	MUL	0.00070	1.17 (1.04–1.31)	0.016
rs3769249	RAPGEF4	2	173,648,169	G/A	0.647	0.0040	0.79	DOM	0.00077	1.27 (1.06–1.51)	0.018
rs17280845	CBLB	3	106,927,226	T/C	0.238	0.00083	0.65	REC	0.010	1.37 (1.07–1.76)	0.027
rs5219‡	KCNH1	11	17,366,148	T/C	0.476	0.0054	0.45	MUL	0.014	1.11 (1.02–1.20)	0.031
rs10915239	MECR	1	29,344,565	C/A	0.945	0.0046	0.60	REC	0.016	1.26 (1.04–1.51)	0.033
rs11206883	PRKAA2	1	56,815,240	A/G	0.095	0.0014	0.58	MUL	0.026	1.17 (1.02–1.34)	0.054
rs11183212	ARID2	12	44,500,134	G/A	0.200	0.00036	0.68	MUL	0.028	1.12 (1.01–1.24)	0.061
rs2395836	PRKAR2B	7	106,381,475	C/T	0.519	0.0022	0.26	DOM	0.034	1.16 (1.01–1.34)	0.072
rs2970871	PPARGC1A	4	23,566,851	C/T	0.424	0.0012	0.081	REC	0.042	1.17 (1.01–1.36)	0.088
rs11888	JAK3	19	17,796,626	C/T	0.315	0.0014	0.71	MUL	0.075	1.08 (0.99–1.18)	0.15
rs2235718	FOXO1	6	1,552,602	T/C	0.117	0.00068	0.28	REC	0.096	1.55 (0.92–2.59)	0.19
rs3118526	RXRA	9	134,563,302	C/T	0.922	0.0039	0.60	DOM	0.11	0.52 (0.23–1.18)	0.21
rs9313	SORBS1	10	97,061,862	G/T	0.919	0.0045	0.66	MUL	0.11	1.13 (0.97–1.32)	0.21
rs9402346	ENPPI	6	132,185,981	C/G	0.646	0.00062	¶	—	—	—	—
rs1830971	ENPPI	6	132,190,046	A/G	0.648	0.00072	¶	—	—	—	—
rs1409184	ENPPI	6	132,182,184	G/A	0.646	0.00072	¶	—	—	—	—
rs6802898	PPARG	3	12,366,207	C/T	0.835	0.0031	¶	—	—	—	—
rs7796553	NRF1	7	128,974,194	C/T	0.172	0.0039	¶	—	—	—	—
rs943852	RAPGEF1	9	131,480,747	T/C	0.111	0.0042	¶	—	—	—	—

\*Positions based on hg17. †SNP was originally imputed (see online appendix Table 4). ‡Nonsynonymous SNP selected for stage 2 genotyping. §Included even though stage 2 sample success rate <90%. ¶SNP was not successfully genotyped in stage 2 or not selected for genotyping in stage 2 based on high LD with a selected SNP.

TABLE 5  
Quantitative trait association results for SNPs genotyped in FUSION stage 1 and 2 samples

SNP	Gene	Chromosome	Position (bp)	Major/minor allele	Trait	Samples*	Stage 1 $P$ value†	Stage 2 $P$ value†	Combined $P$ value‡
rs9615264	<i>PPARA</i>	22	44,953,108	G/A	HDL level	4,682	1.06E-04	0.13	0.00013
rs10517844	<i>CPE</i>	4	166,691,996	T/C	Cholesterol-to	4,682	4.00E-05	0.66	0.009
					HDL ratio				
rs4689388	<i>WFS1</i>	4	6,388,128	A/G	HDL level	4,067	5.30E-05	0.94	0.002
					LDL level				
rs429358	<i>APOE</i>	19	50,103,781	T/C	Cholesterol-to	2,327	1.78E-10	§	—
					HDL ratio				
					LDL level				
rs4912407	<i>PRKAA2</i>	1	56,825,022	G/A	HDL level	2,327	1.51E-05	§	—
					LDL level				
					Triglyceride level				
					Triglyceride-to				
					HDL ratio	2,339	2.77E-05	§	—

\*Number of samples corrected to an effective sample size considering the relatedness of some samples. † $P$  value calculated under additive model. ‡Stage 1 and 2  $P$  values combined by meta-analysis (see RESEARCH DESIGN AND METHODS). §SNP was not successfully genotyped in stage 2.

diabetes in the stage 1 and 2 samples was SNP rs4740283 ( $P_{\text{SNP}} = 0.00013$ ), located 4 kb downstream of Rap guanine nucleotide exchange factor 1 (*RAPGEF1*). *RAPGEF1* is a ubiquitously expressed gene involved in insulin signaling (41) and Ras-mediated tumor suppression (42). rs4740283 is in strong LD with SNPs in the coding region and may affect either a regulatory element or protein function. Variation in this gene may contribute to susceptibility through reduced ability of peripheral tissues to absorb glucose in response to insulin.

The second strongest-associated SNP in the stage 1 and 2 samples was Arg72Pro in *TP53* (rs1042522,  $P_{\text{SNP}} = 0.00086$ ), which was originally identified by imputation, subsequently genotyped, and not well tagged by any originally genotyped SNP (maximum  $r^2 = 0.27$  with rs2909430). *TP53* encodes the tumor suppressor protein p53, and the Arg72Pro variant has a functional role in the efficiency of p53 in inducing apoptosis, possibly through reduced localization to the mitochondria (43). The risk allele Arg72 has higher apoptotic potential, which is consistent with a possible link between increased pancreatic  $\beta$ -cell apoptosis, impaired insulin secretion, and type 2 diabetes.

We observed significant association with SNPs in two genes previously implicated in type 2 diabetes susceptibility, nuclear respiratory factor 1 (*NRF1*) and the insulin-dependent facilitated glucose transporter *SLC2A2*. *NRF1* helps regulate mitochondrial transcription and oxidative phosphorylation (44), which has a known role in insulin resistance, and the associated *NRF1* variant, rs1882095, is located 1 kb downstream of the gene and not in modest LD ( $r^2 > 0.6$ ) with any HapMap SNP. In *SLC2A2* we found supporting evidence in stage 1 for the nonsynonymous variant Thr110Ile (rs5400) ( $P_{\text{SNP}} = 0.0065$ ), as well as a previously unreported variant, rs10513684 ( $P_{\text{SNP}} = 0.0046$ ). The rs10513684 signal became slightly more significant after stage 2 genotyping ( $P_{\text{SNP}} = 0.0023$ ); however, the signal was attenuated ( $P = 0.18$ ) after inclusion of Thr110Ile in the analysis.

Among the most significant type 2 diabetes-associated SNPs is rs2021966 in *ENPP1* ( $P_{\text{SNP}} = 0.00026$ ). SNPs in high LD with rs2021966 are located in intron 1, in a region of strong multispecies conservation containing a pseudo-

gene but no known transcripts. Previous studies of *ENPP1* have reported associations with rs1044498 and with a related three-SNP haplotype (rs1044498, rs1799774, and rs7754561) and support a modest role in type 2 diabetes susceptibility, possibly acting through obesity (45). In our study, rs1044498 ( $P_{\text{SNP}} = 0.16$ ) and rs7754561 ( $P_{\text{SNP}} = 0.18$ ,  $r^2 = 1$  with rs7754561) were not significantly associated with type 2 diabetes (rs1799774 was not tested). The newly identified variants are in very low LD with rs1044498 ( $r^2 < 0.05$ ).

Although we observed significant quantitative trait associations in previously implicated genes (*APOE* and *PPARA* with serum lipid levels), no quantitative trait associations became more significant after addition of stage 2 samples (Table 5). This is likely due in part to the small number of SNPs selected for follow-up. Stage 2 genotyping of SNPs less significant in stage 1 samples will be necessary to establish whether any novel SNPs contribute to quantitative trait variability.

In any gene-based study, the definition of gene boundaries is critical but, by necessity, somewhat arbitrary. We defined a gene region as 10 kb upstream of the first known exon through 5 kb downstream of the last known exon in an attempt to capture the majority of nearby regulatory elements influencing a gene. Regulatory elements, however, can often be found up to several hundred kilobases away from a gene (46). We evaluated whether a broader definition of a gene had a substantial effect on the  $P_{\text{gene}}$  results by testing extended gene regions 50 kb upstream and 50 kb downstream of transcribed regions and by including HumanHap300 SNPs from these regions in our analysis. Using the extended gene boundaries, the insulin gene *INS* would be the most significant gene in our study ( $P_{\text{gene}} = 0.0019$ ), driven by SNP rs10743152 ( $P_{\text{SNP}} = 0.00015$ ) located 13 kb upstream of the first exon. Other genes that had significant SNPs ( $P_{\text{gene}} < 0.05$ ) only in the extended gene region were *MAP2K1*, *CDK4*, and *IRF4*.

Even using the narrow gene boundaries, several SNPs in our study may influence expression or function of other nearby or even more distant genes. Recent GWA studies have confirmed novel susceptibility variants

downstream of *HHEX*, a gene selected for this study by CAESAR (9–12); the reported SNPs are located outside of the narrow gene region (–10 kb/+5 kb) in a large LD block that includes *KIF11* and *IDE*, and we only detected nominal significance in the narrow *HHEX* region ( $P_{\text{SNP}} = 0.037$  for rs12262390). For some genes, the extent of LD surrounding significant SNPs implicates flanking genes. For example, in *ARID2*, rs35115 ( $P_{\text{SNP}} = 0.0067$ ) is located in intron 7 but also tags the nonsynonymous variant rs7315731 in *SFRS2IP* ( $r^2 = 0.93$ ). These examples demonstrate that defining a gene boundary requires a balance between capturing all possible SNPs influencing the gene and introducing SNPs that may be more functionally relevant to other genes. A more sophisticated approach to establish gene boundaries that defines each gene boundary separately by considering the genomic context around the gene may be helpful in future gene-based approaches.

Gene-based approaches to interpreting the results of candidate gene and even genome-wide association studies are important because most variation influencing susceptibility to type 2 diabetes and other common complex traits is currently expected to be gene centric, although the definition of a gene is constantly evolving. Detailed coverage of the common variation in these genes represents a critical requirement for an effective and thorough gene-based study. Here, we have identified genes significantly associated with type 2 diabetes and related quantitative traits that are attractive targets for future replication studies. Confirmation in a larger sample set and meta-analyses across studies will be important to help determine the role of these genes.

#### ACKNOWLEDGMENTS

Support for this research was provided by National Institutes of Health (NIH) Grants DK072193 (to K.L.M.) and DK062370 (to M.B.), a postdoctoral fellowship award from the American Diabetes Association (to C.J.W.), and the National Center for Integrative Biomedical Informatics (NCIBI) at the University of Michigan (U54 DA021519). K.L.M. and G.R.A. are Pew Scholars in the Biomedical Sciences. Genome-wide genotyping was performed by the Johns Hopkins University Genetic Resources Core Facility (GRCF) SNP Center at the Center for Inherited Disease Research (CIDR), with support from CIDR NIH contract no. N01-HG-65403 and the GRCF SNP Center.

We thank the Finnish citizens who generously participated in this study, Michael Andre and Rachana Kshatriya of the University of North Carolina Mammalian Genotyping Core for Illumina GoldenGate genotyping, Amy Swift and Mario Morken of the NHGRI for stage 2 genotyping, and Kurt Hetrick, Michael Barnhart, Craig Bark, Janet Goldstein, and Lee Watkins of the CIDR for expert technical work on genome-wide Illumina Infinium genotyping.

#### REFERENCES

- Kaprio J, Tuomilehto J, Koskenvuo M, Romanov K, Reunanen A, Eriksson J, Stengård J, Kesäniemi YA: Concordance for type 1 (insulin-dependent) and type 2 (non-insulin-dependent) diabetes mellitus in a population-based cohort of twins in Finland. *Diabetologia* 35:1060–1067, 1992
- Frayling TM: Genome-wide association studies provide new insights into type 2 diabetes aetiology. *Nat Rev Genet* 8:657–662, 2007
- Freeman H, Cox RD: Type-2 diabetes: a cocktail of genetic discovery. *Hum Mol Genet* 15:R202–R209, 2006
- Barroso I, Luan J, Middelberg RP, Harding AH, Franks PW, Jakes RW, Clayton D, Schafer AJ, O'Rahilly S, Wareham NJ: Candidate gene association study in type 2 diabetes indicates a role for genes involved in beta-cell function as well as insulin action. *PLoS Biol* 1:E20, 2003
- Willer CJ, Bonnycastle LL, Conneely KN, Duren WL, Jackson AU, Scott LJ, Narisu N, Chines PS, Skol A, Stringham HM, Petrie J, Erdos MR, Swift AJ, Enloe ST, Sprau AG, Smith E, Tong M, Doheny KF, Pugh EW, Watanabe RM, Buchanan TA, Valle TT, Bergman RN, Tuomilehto J, Mohlke KL, Collins FS, Boehnke M: Screening of 134 single nucleotide polymorphisms (SNPs) previously associated with type 2 diabetes replicates association with 12 SNPs in nine genes. *Diabetes* 56:256–264, 2007
- Altshuler D, Hirschhorn JN, Klannemark M, Lindgren CM, Vohl MC, Nemesh J, Lane CR, Schaffner SF, Bolk S, Brewer C, Tuomi T, Gaudet D, Hudson TJ, Daly M, Groop L, Lander ES: The common PPARGgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat Genet* 26:76–80, 2000
- Gloyn AL, Weedon MN, Owen KR, Turner MJ, Knight BA, Hitman G, Walker M, Levy JC, Sampson M, Halford S, McCarthy MI, Hattersley AT, Frayling TM: Large-scale association studies of variants in genes encoding the pancreatic  $\beta$ -cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with type 2 diabetes. *Diabetes* 52:568–572, 2003
- Grant SF, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J, Helgason A, Stefansson H, Emilsson V, Helgadóttir A, Styrkarsdóttir U, Magnusson KP, Walters GB, Palsdóttir E, Jonsdóttir T, Gudmundsdóttir T, Gylfason A, Saemundsdóttir J, Wilensky RL, Reilly MP, Rader DJ, Bagger Y, Christiansen C, Gudnason V, Sigurdsson G, Thorsteinsdóttir U, Gulcher JR, Kong A, Stefansson K: Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat Genet* 38:320–323, 2006
- Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S, Balkau B, Heude B, Charpentier G, Hudson TJ, Montpetit A, Pshezhetsky AV, Prentki M, Posner BI, Balding DJ, Meyre D, Polychronakos C, Froguel P: A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 445:881–885, 2007
- Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-Olsson L, Ding CJ, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li XY, Conneely KN, Riebow NL, Sprau AG, Tong M, White PP, Hetrick KN, Barnhart MW, Bark CW, Goldstein JL, Watkins L, Xiang F, Saramies J, Buchanan TA, Watanabe RM, Valle TT, Kinnunen L, Abecasis GR, Pugh EW, Doheny KF, Bergman RN, Tuomilehto J, Collins FS, Boehnke M: A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 316:1341–1345, 2007
- Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, Timpson NJ, Perry JR, Rayner NW, Freathy RM, Barrett JC, Shields B, Morris AP, Ellard S, Groves CJ, Harries LW, Marchini JL, Owen KR, Knight B, Cardon LR, Walker M, Hitman GA, Morris AD, Doney A, the Wellcome Trust Case Control Consortium (WTCCC), McCarthy MI, Hattersley AT: Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 316:1336–1341, 2007
- Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research, Saxena R, Voight BF, Lyssenko V, Burt NP, de Bakker PI, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Daly MJ, Hughes TE, Groop L, Altshuler D, Almgren P, Florez JC, Meyer J, Ardlie K, Bengtsson Boström K, Isomaa B, Lettre G, Lindblad U, Lyon HN, Melander O, Newton-Cheh C, Nilsson P, Orho-Melander M, Rastam L, Speliotes EK, Taskinen MR, Tuomi T, Guiducci C, Berglund A, Carlson J, Gianniny L, Hackett R, Hall L, Holmkvist J, Laurila E, Sjogren M, Sterner M, Surti A, Svensson M, Svensson M, Tewhey R, Blumensiel B, Parkin M, Defelice M, Barry R, Brodeur W, Camarata J, Chia N, Fava M, Gibbons J, Handsaker B, Healy C, Nguyen K, Gates C, Sougnez C, Gage D, Nizzari M, Gabriel SB, Chirn GW, Ma Q, Parikh H, Richardson D, Rieke D, Purcell S: Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 316:1331–1336, 2007
- Steinthorsdóttir V, Thorleifsson G, Reynisdóttir I, Benediktsson R, Jonsdóttir T, Walters GB, Styrkarsdóttir U, Gretarsdóttir S, Emilsson V, Ghosh S, Baker A, Snorrardóttir S, Bjarnason H, Ng MC, Hansen T, Bagger Y, Wilensky RL, Reilly MP, Adeyemo A, Chen Y, Zhou J, Gudnason V, Chen G, Huang H, Lashley K, Doumatey A, So WY, Ma RC, Andersen G, Borch-Johnsen K, Jorgensen T, van Vliet-Ostapchouk JV, Hofker MH, Wijmenga C, Christiansen C, Rader DJ, Rotimi C, Gurney M, Chan JC, Pedersen O, Sigurdsson G, Gulcher JR, Thorsteinsdóttir U, Kong A, Stefansson K: A variant in *CDKAL1* influences insulin response and risk of type 2 diabetes. *Nat Genet* 39:770–775, 2007
- Salonen JT, Uimari P, Aalto JM, Pirskanen M, Kaikkonen J, Todorova B, Hyppönen J, Korhonen VP, Asikainen J, Devine C, Tuomainen TP, Luode-



- mann J, Nauck M, Kerner W, Stephens RH, New JP, Ollier WE, Gibson JM, Payton A, Horan MA, Pendleton N, Mahoney W, Meyre D, Delplanque J, Froguel P, Luzzatto O, Yakir B, Darvasi A: Type 2 diabetes whole-genome association study in four populations: the DiaGen consortium. *Am J Hum Genet* 81:338–345, 2007
15. Hayes MG, Pluzhnikov A, Miyake K, Sun Y, Ng MC, Roe CA, Below JE, Nicolae RI, Konkashbaev A, Bell GI, Cox NJ, Hanis CL: Identification of type 2 diabetes genes in Mexican Americans through genome-wide association studies. *Diabetes* 56:3033–3044, 2007
  16. Florez JC, Manning AK, Dupuis J, McAteer J, Irenze K, Gianniny L, Mirel DB, Fox CS, Cupples LA, Meigs JB: A 100K genome-wide association scan for diabetes and related traits in the Framingham Heart Study: replication and integration with other genome-wide datasets. *Diabetes* 56:3063–3074, 2007
  17. Hanson RL, Bogardus C, Duggan D, Kobes S, Knowlton M, Infante AM, Marovich L, Benitez D, Baier LJ, Knowler WC: A search for variants associated with young-onset type 2 diabetes in American Indians in a 100K genotyping array. *Diabetes* 56:3045–3052, 2007
  18. Rampersaud E, Damcott CM, Fu M, Shen H, McArdle P, Shi X, Shelton J, Yin J, Chang CY, Ott SH, Zhang L, Zhao Y, Mitchell BD, O'connell J, Shuldiner AR: Identification of novel candidate genes for type 2 diabetes from a genome-wide association scan in the Old Order Amish: evidence for replication from diabetes-related quantitative traits and from independent populations. *Diabetes* 56:3053–3062, 2007
  19. Sandhu MS, Weedon MN, Fawcett KA, Wasson J, Debenham SL, Daly A, Lango H, Frayling TM, Neumann RJ, Sherva R, Blech I, Pharoah PD, Palmer CN, Kimber C, Tavendale R, Morris AD, McCarthy MI, Walker M, Hitman G, Glaser B, Permutt MA, Hattersley AT, Wareham NJ, Barroso I: Common variants in WFS1 confer risk of type 2 diabetes. *Nat Genet* 39:951–953, 2007
  20. Gudmundsson J, Sulem P, Steinthorsdottir V, Bergthorsson JT, Thorleifsson G, Manolescu A, Rafnar T, Gudbjartsson D, Agnarsson BA, Baker A, Sigurdsson A, Benediktsson KR, Jakobsdottir M, Blondal T, Stacey SN, Helgason A, Gunnarsdottir S, Olafsdottir A, Kristinsson KT, Birgisdottir B, Ghosh S, Thorlacius S, Magnusdottir D, Stefansdottir G, Kristjansson K, Bagger Y, Wilensky RL, Reilly MP, Morris AD, Kimber CH, Adeyemo A, Chen Y, Zhou J, So WY, Tong PC, Ng MC, Hansen T, Andersen G, Borch-Johnsen K, Jorgensen T, Tres A, Fuertes F, Ruiz-Echarrri M, Asin L, Saez B, van Boven E, Klaver S, Swinkels DW, Aben KK, Graif T, Cashy J, Suarez BK, van Viesssen Trip O, Frigge ML, Ober C, Hofker MH, Wijmenga C, Christiansen C, Rader DJ, Palmer CN, Rotimi C, Chan JC, Pedersen O, Sigurdsson G, Benediktsson R, Jonsson E, Einarsson GV, Mayordomo JJ, Catalona WJ, Kiemeny LA, Barkardottir RB, Gulcher JR, Thorsteinsdottir U, Kong A, Stefansson K: Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat Genet* 39:977–983, 2007
  21. Winckler W, Weedon MN, Graham RR, McCarroll SA, Purcell S, Almgren P, Tuomi T, Gaudet D, Bosström KB, Walker M, Hitman G, Hattersley AT, McCarthy MI, Ardlie KG, Hirschhorn JN, Daly MJ, Frayling TM, Groop L, Altshuler D: Evaluation of common variants in the six known maturity-onset diabetes of the young (MODY) genes for association with type 2 diabetes. *Diabetes* 56:685–693, 2007
  22. Valle T, Tuomilehto J, Bergman RN, Ghosh S, Hauser ER, Eriksson J, Nylund SJ, Kohtamaki K, Toivanen L, Vidgren G, Tuomilehto-Wolf E, Ehnholm C, Blaschak J, Langefeld CD, Watanabe RM, Magnusson V, Ally DS, Hagopian WA, Ross E, Buchanan TA, Collins F, Boehnke M: Mapping genes for NIDDM: design of the Finland-United States Investigation of NIDDM Genetics (FUSION) study. *Diabetes Care* 21:949–958, 1998
  23. Silander K, Mohlke KL, Scott LJ, Peck EC, Hollstein P, Skol AD, Jackson AU, Deloukas P, Hunt S, Stavrides G, Chines PS, Erdos MR, Narisu N, Conneely KN, Li C, Fingerlin TE, Dhanjal SK, Valle TT, Bergman RN, Tuomilehto J, Watanabe RM, Boehnke M, Collins FS: Genetic variation near the hepatocyte nuclear factor-4  $\alpha$  gene predicts susceptibility to type 2 diabetes. *Diabetes* 53:1141–1149, 2004
  24. Bonnycastle LL, Willer CJ, Conneely KN, Jackson AU, Burrill CP, Watanabe RM, Chines PS, Narisu N, Scott LJ, Enloe ST, Swift AJ, Duren WL, Stringham HM, Erdos MR, Riebow NL, Buchanan TA, Valle TT, Tuomilehto J, Bergman RN, Mohlke KL, Boehnke M, Collins FS: Common variants in maturity-onset diabetes of the young genes contribute to risk of type 2 diabetes in Finns. *Diabetes* 55:2534–2540, 2006
  25. Scott LJ, Bonnycastle LL, Willer CJ, Sprau AG, Jackson AU, Narisu N, Duren WL, Chines PS, Stringham HM, Erdos MR, Valle TT, Tuomilehto J, Bergman RN, Mohlke KL, Collins FS, Boehnke M: Association of transcription factor 7-like 2 (TCF7L2) variants with type 2 diabetes in a Finnish sample. *Diabetes* 55:2649–2653, 2006
  26. Gaulton KJ, Mohlke KL, Vision TJ: A computational system to select candidate genes for complex human traits. *Bioinformatics* 23:1132–1140, 2007
  27. International HapMap Consortium: A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449:851–861, 2007
  28. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR: Markov model for rapid haplotyping and genotype imputation in genome wide studies. Submitted, 2007
  29. Silander K, Scott LJ, Valle TT, Mohlke KL, Stringham HM, Wiles KR, Duren WL, Doheny KF, Pugh EW, Chines P, Narisu N, White PP, Fingerlin TE, Jackson AU, Li C, Ghosh S, Magnusson VL, Colby K, Erdos MR, Hill JE, Hollstein P, Humphreys KM, Kasad RA, Lambert J, Lazaridis KN, Lin G, Morales-Mena A, Patzkowski K, Pfahl C, Porter R, Rha D, Segal L, Suh YD, Tovar J, Unni A, Welch C, Douglas JA, Epstein MP, Hauser ER, Hagopian W, Buchanan TA, Watanabe RM, Bergman RN, Tuomilehto J, Collins FS, Boehnke M, the Finland-United States Investigation of NIDDM Genetics (FUSION): A large set of Finnish affected sibling pair families with type 2 diabetes suggests susceptibility loci on chromosomes 6, 11, and 14. *Diabetes* 53:821–829, 2004
  30. Saaristo T, Peltonen M, Lindstrom J, Saarikoski L, Sundvall J, Eriksson JG, Tuomilehto J: Cross-sectional evaluation of the Finnish diabetes risk score: a tool to identify undetected type 2 diabetes, abnormal glucose tolerance and metabolic syndrome. *Diab Vasc Dis Res* 2:67–72, 2005
  31. Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C, Richter J, Rubin GM, Blake JA, Bult C, Dolan M, Drabkin H, Eppig JT, Hill DP, Ni L, Ringwald M, Balakrishnan R, Cherry JM, Christie KR, Costanzo MC, Dwight SS, Engel S, Fisk DG, Hirschman JE, Hong EL, Nash RS, Sethuraman A, Theesfeld CL, Botstein D, Dolinski K, Feierbach B, Berardini T, Mundodi S, Rhee SY, Apweiler R, Barrell D, Camon E, Dimmer E, Lee V, Chisholm R, Gaudet P, Kibbe W, Kishore R, Schwarz EM, Sternberg P, Gwinn M, Hannick L, Wortman J, Berriman M, Wood V, de la Cruz N, Tonellato P, Jaiswal P, Seigfried T, White R, the Gene Ontology Consortium: The gene ontology (GO) database and informatics resource. *Nucleic Acid Res* 32:D258–D261, 2004
  32. Kelso J, Visagie J, Theiler G, Christoffels A, Barden S, Smedley D, Otgaar D, Greyling G, Jongeneel CV, McCarthy MI, Hide T, Hide W: eVOC: A controlled vocabulary for unifying gene expression data. *Genome Res* 13:1222–1230, 2003
  33. Smith CL, Goldsmith CA, Eppig JT: The mammalian phenotype ontology as a tool for annotating, analyzing and comparing phenotypic information. *Genome Biol* 6:R7, 2005
  34. Fingerlin TE, Erdos MR, Watanabe RM, Wiles KR, Stringham HM, Mohlke KL, Silander K, Valle TT, Buchanan TA, Tuomilehto J, Bergman RN, Boehnke M, Collins FS: Variation in three single nucleotide polymorphisms in the calpain-10 gene not associated with type 2 diabetes in a large Finnish cohort. *Diabetes* 51:1644–1648, 2002
  35. Horikawa Y, Oda N, Cox NJ, Li X, Orho-Melander M, Hara M, Hinokio Y, Lindner TH, Mashima H, Schwarz PE, del Bosque-Plata L, Horikawa Y, Oda Y, Yoshiuchi I, Colilla S, Polonsky KS, Wei S, Concannon P, Iwasaki N, Schulze J, Baier LJ, Bogardus C, Groop L, Boerwinkle E, Hanis CL, Bell GI: Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat Genet* 26:163–175, 2000
  36. Weedon MN, Schwarz PE, Horikawa Y, Iwasaki N, Illig T, Holle R, Rathmann W, Selisko T, Schulze J, Owen KR, Evans J, Del Bosque-Plata L, Hitman G, Walker M, Levy JC, Sampson M, Bell GI, McCarthy MI, Hattersley AT, Frayling TM: Meta-analysis and a large association study confirm a role for calpain-10 variation in type 2 diabetes susceptibility. *Am J Hum Genet* 73:1208–1212, 2003
  37. Conneely KN, Boehnke M: So many correlated tests, so little time! Rapid adjustment of p-values for multiple correlated tests. *Am J Hum Genet* 81:1158–1168, 2007
  38. Chen WM, Abecasis GR: Family-based association tests for genomewide association scans. *Am J Hum Genet* 81:913–926, 2007
  39. Tai ES, Demissie S, Cupples LA, Corella D, Wilson PW, Schaefer EJ, Ordovas JM: Association between the PPARA L162V polymorphism and plasma lipid levels: the Framingham Offspring Study. *Arterioscler Thromb Vasc Biol* 22:805–810, 2002
  40. Knoblauch H, Bauerfeind A, Krahenbuhl C, Dauray A, Rohde K, Bejanin S, Essioux L, Schuster H, Luft FC, Reich JG: Common haplotypes in five genes influence genetic variance of LDL and HDL cholesterol in the general population. *Hum Mol Genet* 11:1477–1485, 2002
  41. Chiang SH, Chang L, Saitli AR: TC10 and insulin-stimulated glucose transport. *Methods Enzymol* 406:701–714, 2006
  42. Guerrero C, Martín-Encabo S, Fernández-Medarde A, Santos E: C3G-mediated suppression of oncogene-induced focus formation in fibroblasts involves inhibition of ERK activation, cyclin A expression and alterations of anchorage-independent growth. *Oncogene* 23:4885–4893, 2004

43. Dumont P, Leu JI, Della Pietra AC 3rd, George DL, Murphy M: The codon 72 polymorphic variants of p53 have markedly different apoptotic potential. *Nat Genet* 33:357–365, 2003
44. Patti ME, Butte AJ, Crunkhorn S, Cusi K, Berria R, Kashyap S, Miyazaki Y, Kohane I, Costello M, Saccone R, Landaker EJ, Goldfine AB, Mun E, DeFronzo R, Finlayson J, Kahn CR, Mandarino LJ: Coordinated reduction of genes of oxidative metabolism in humans with insulin resistance and diabetes: potential role of PGC1 and NRF1. *Proc Natl Acad Sci U S A* 100:8466–8471, 2003
45. Meyre D, Bouatia-Naji N, Tounian A, Samson C, Lecoeur C, Vatin V, Ghossaini M, Wachter C, Hercberg S, Charpentier G, Patsch W, Pattou F, Charles MA, Tounian P, Clement K, Jouret B, Weill J, Maddux BA, Goldfine ID, Walley A, Boutin P, Dina C, Froguel P: Variants of ENPP1 are associated with childhood and adult obesity and increase the risk of glucose intolerance and type 2 diabetes. *Nat Genet* 37:863–867, 2005
46. Bondarenko VA, Liu YV, Jiang YI, Studitsky VM: Communication over a large distance: enhancers and insulators. *Biochem Cell Biol* 81:241–251, 2003