

Interrogating Type 2 Diabetes Genome-Wide Association Data Using a Biological Pathway-Based Approach

John R.B. Perry,¹ Mark I. McCarthy,^{2,3} Andrew T. Hattersley,¹ Eleftheria Zeggini,² the Wellcome Trust Case Control Consortium, Michael N. Weedon,¹ and Timothy M. Frayling¹

OBJECTIVE—Recent genome-wide association studies have resulted in a dramatic increase in our knowledge of the genetic loci involved in type 2 diabetes. In a complementary approach to these single-marker studies, we attempted to identify biological pathways associated with type 2 diabetes. This approach could allow us to identify additional risk loci.

RESEARCH DESIGN AND METHODS—We used individual level genotype data generated from the Wellcome Trust Case Control Consortium (WTCCC) type 2 diabetes study, consisting of 393,143 autosomal SNPs, genotyped across 1,924 case subjects and 2,938 control subjects. We sought additional evidence from summary level data available from the Diabetes Genetics Initiative (DGI) and the Finland-United States Investigation of NIDDM Genetics (FUSION) studies. Statistical analysis of pathways was performed using a modification of the Gene Set Enrichment Algorithm (GSEA). A total of 439 pathways were analyzed from the Kyoto Encyclopedia of Genes and Genomes, Gene Ontology, and BioCarta databases.

RESULTS—After correcting for the number of pathways tested, we found no strong evidence for any pathway showing association with type 2 diabetes (top $P_{\text{adj}} = 0.31$). The candidate WNT-signaling pathway ranked top (nominal $P = 0.0007$, excluding TCF7L2; $P = 0.002$), containing a number of promising single gene associations. These include CCND2 (rs11833537; $P = 0.003$), SMAD3 (rs7178347; $P = 0.0006$), and PRICKLE1 (rs1796390; $P = 0.001$), all expressed in the pancreas.

CONCLUSIONS—Common variants involved in type 2 diabetes risk are likely to occur in or near genes in multiple pathways. Pathway-based approaches to genome-wide association data may be more successful for some complex traits than others, depending on the nature of the underlying disease physiology. *Diabetes* 58:1463–1467, 2009

Recent genome-wide association (GWA) studies have resulted in a dramatic increase in our knowledge of the genetic loci involved in type 2 diabetes. There are now 18 common variants with robust evidence of association with the disease (1–7). Individually, these variants only have a small effect on disease risk, and in combination they account for only

~3% of the heritability of type 2 diabetes (6). GWA studies have focused on only the most statistically strongly associated individual variants, with robust association being declared only if a variant meets a stringent “genome-wide” level of statistical significance. However, GWA studies of a few thousand individuals are underpowered to detect variants with odds ratios of <1.2 at $P < 1 \times 10^{-5}$ —the typical P values used to take SNPs forward into replication studies. This means many real type 2 diabetes predisposing variants may not be detected in GWA data and are not yet followed up. The most robust way of testing for these variants is to increase sample sizes, and such efforts are ongoing. However, ever larger numbers of case and control subjects may be difficult to obtain. We therefore sought alternative approaches to investigating GWA data.

RESEARCH DESIGN AND METHODS

Samples. For the initial analysis, we used individual-level genotype data generated from the Wellcome Trust Case Control Consortium (WTCCC) type 2 diabetes study. Quality-control criteria and study design are described in detail in Zeggini et al. (7). We used a total of 393,143 autosomal single nucleotide polymorphisms (SNPs), in 1,924 case subjects and 2,938 population-based control subjects.

We used summary association statistics from two type 2 diabetes GWA studies for validation. These are the Diabetes Genetics Initiative (DGI) and the Finland-United States Investigation of NIDDM Genetics (FUSION) study, which together consist of 2,625 case subjects and 2,641 control subjects (2,3,6). Imputed SNP association statistics were used for any FUSION or DGI SNPs that were missing or not directly genotyped. Only those with high imputation confidence ($r^2 \text{ hat} > 0.4$) were included in the analysis.

Pathway databases. The Gene Set Enrichment Algorithm (GSEA) modified algorithm was provided as part of the GenGen suite (<http://openbioinformatics.org/gengen>), complete with formatted data files using the SNP/Gene mappings from the UCSC Genome Browser annotation in the snp126, refGene, and refLink tables (genome build:hg18). The software uses human biological pathways defined from Gene Ontology (<http://www.geneontology.org>), BioCarta (<http://www.biocarta.com>), and the Kyoto Encyclopedia of Genes and Genomes (KEGG; <http://www.genome.jp/kegg>) databases (8,9). All human KEGG and BioCarta pathways were used in the analysis, but only level 4 annotations in Biological Process and Molecular Function from Gene Ontology. These were the latest version when downloaded from the Gengen website (<http://openbioinformatics.org/gengen>, accessed December 2007).

We also designed three artificial positive control pathways. The first contained 17 known type 2 diabetes loci (18 confirmed variants excluding SLC30A8 not captured on the Affymetrix chip), and the second and third contained the same genes, diluted with 17 and 34 genes with median GWA χ^2 association statistics.

Bioinformatic and statistical analysis. Analysis of pathways was performed using a recently published algorithm (10). This method implements the existing GSEA algorithm (11). Briefly, χ^2 disease association statistics from all quality-controlled SNPs were mapped onto genes. The mapping process assigns the SNP with the highest test statistic, within a 200-kb flanking window on either side of the gene, to represent the overall test statistic for an individual gene. SNPs that are not mapped to any defined gene are discarded from further analysis, whereas SNPs within the region of multiple genes are mapped to all. This produced a list of 303,861 SNPs in 15,114 genes. Each of these 15,114 genes is then ranked from highest to lowest according to their test statistic (represented by the highest associated SNP). From an initial set of 2,077 KEGG, Gene Ontology, and BioCarta pathways, 439 pathways met the

From ¹Genetics of Complex Traits, Institute of Biomedical and Clinical Science, Peninsula Medical School, Magdalen Road, Exeter, U.K.; the ²Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, U.K.; and the ³Oxford Centre for Diabetes, Endocrinology and Medicine, University of Oxford, Churchill Hospital, Oxford, U.K.

Corresponding author: Michael Weedon, michael.weedon@pms.ac.uk.

Received 7 October 2008 and accepted 18 February 2009.

Published ahead of print at <http://diabetes.diabetesjournals.org> on 27 February 2009. DOI: 10.2337/db08-1378.

© 2009 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.

TABLE 1
All pathways with nominal $P \leq 0.05$ in WTCCC data

Database	Pathway	Size	P (P_{adj})	False discovery rate
KEGG	WNT signaling pathway	126	0.0007 (0.31)	0.20
KEGG	Olfactory transduction	26	0.0009 (0.4)	0.37
GO	Organic acid biosynthetic process	30	0.004 (1)	0.60
GO	Regulation of Wnt receptor signaling pathway	24	0.005 (1)	0.51
GO	Odontogenesis	22	0.005 (1)	0.54
GO	Aminoglycan metabolic process	32	0.007 (1)	0.53
GO	Membrane invagination	62	0.009 (1)	0.51
GO	Calcium-dependent cell-cell adhesion	20	0.01 (1)	0.60
KEGG	Galactose metabolism	26	0.01 (1)	0.53
BioCarta	ALK in cardiac myocytes	32	0.01 (1)	0.46
GO	Biom mineral formation	30	0.02 (1)	0.64
GO	Endocytosis	131	0.02 (1)	0.64
GO	Gonad development	22	0.02 (1)	0.63
GO	Sensory organ development	59	0.03 (1)	0.74
BioCarta	WNT signaling pathway	20	0.03 (1)	0.72
KEGG	Pyruvate metabolism	38	0.03 (1)	0.76
GO	Monosaccharide metabolic process	115	0.03 (1)	0.78
GO	Cysteine-type peptidase activity	114	0.03 (1)	0.77
GO	Carbohydrate biosynthetic process	73	0.04 (1)	0.82
GO	GABA receptor activity	20	0.04 (1)	0.81
KEGG	Type II diabetes	36	0.04 (1)	0.82
GO	Positive regulation of transport	26	0.04 (1)	0.78
KEGG	TGF- β signaling pathway	73	0.04 (1)	0.79
GO	Epidermal cell differentiation	45	0.05 (1)	0.75
GO	Epidermis morphogenesis	45	0.05 (1)	0.75
BioCarta	Regulation of PGC-1 α HDAC5	21	0.05 (1)	0.82

Results based on 10,000 permutations and 439 pathways. Gene size is the number of analyzed genes within that pathway. P_{adj} is the Bonferroni-adjusted P value based on 439 statistical tests. GO, Gene Ontology.

criteria of containing between 20 and 200 GWA-captured genes. This cutoff was selected as the optimal range for gene sets based on the power of the algorithm. However, cutoffs of between 10 and 400 produced similar results, with WNT-signaling ranking top.

The significance of each pathway is then judged using a Kolmogorov-Smirnov-like running sum statistic. This test reflects the overrepresentation of pathway genes toward the top of the overall ranked gene list. An enrichment score (ES) is produced, measuring the deviation of the association statistics in a given gene set, compared with a set of randomly picked genes. As larger genes will by chance harbor SNPs with higher test statistics, a normalized enrichment score (NES) is also produced, adjusting for gene size through permutation-based label swapping. By subtracting the mean permuted ES score from the observed ES score and dividing by the SD of the permuted ES scores, different-sized genes and gene sets are directly comparable. A total of 10,000 permutations were performed, where the case/control ratio was maintained during label swapping. The permutation approach gives an empirical P value for an individual pathway based on label swapping, alongside a false discovery rate calculated using NES scores from all permuted values in all pathways examined in an experiment. Full details of formulae used in this algorithm are provided by Wang et al. (10).

RESEARCH DESIGN AND METHODS

Biological pathway-based analysis is a complementary approach to single-point GWA studies. The approach allows us to test the hypothesis that a set of polymorphisms that occur in genes from a particular biological pathway or process are associated with type 2 diabetes. Using GWA data to test for association of a pathway rather than individual genes may provide important novel biological insights (i.e., a pathway associated with type 2 diabetes), which have been missed by the focus on only the most associated individual variants. A pathway-based approach may also help prioritize individual variants, which are at lower levels of statistical confidence, for follow-up in large replication efforts.

Pathway-based approaches have been used extensively for analyzing gene expression data. For example, Mootha et al. (12) used the GSEA approach to identify a set of genes involved in oxidative phosphorylation, whose expression was coordinately decreased in muscle from diabetic patients compared with control subjects. Wang et al. (10) recently published a modified version of the GSEA, which is designed to analyze genome-wide SNP association statistic data rather than expression data. No GWA pathway-based analyses

have yet been performed for type 2 diabetes. In this study, we attempted to identify pathways associated with type 2 diabetes by applying the modified GSEA method to a type 2 diabetes GWA dataset.

We used individual level GWA data from the WTCCC type 2 diabetes study (1,7). We used a total of 393,143 autosomal SNPs genotyped across 1,924 case subjects and 2,938 population-based control subjects. We analyzed 439 pathways defined by the Gene Ontology (8), BioCarta, and KEGG (9) databases. Details of the algorithm used are provided in RESEARCH DESIGN AND METHODS, but briefly, the most strongly associated SNP from a gene ± 200 kb of flanking sequence either side is used to represent each gene in the genome. Overlap is possible, where the same SNP can be mapped to multiple genes. An individual gene set is then assigned a statistic based on the trend for its genes to appear toward the top of the ranked association list of all genes in the genome.

RESULTS

A total of 26 pathways reached a nominal $P < 0.05$, slightly higher than the number expected by chance ($0.05 \times 439 = 22$) (Table 1, Fig. 1). No pathways were associated with type 2 diabetes at $P < 0.05$ after Bonferroni adjustment (top $P = 0.31$). A quantile quantile (QQ) plot of the 439 pathways is shown in Fig. 1. The WNT-signaling pathway was the most strongly associated and was of interest because it contains *TCF7L2*, a common variation that shows the strongest association with type 2 diabetes in all type 2 diabetes GWA studies (7). Excluding the *TCF7L2* gene from analysis ranked the WNT-signaling pathway second out of the 439 pathways (nominal $P = 0.002$). WNT signaling has been proposed as an important candidate pathway for type 2 diabetes, not only because of the *TCF7L2* association but because of its potential importance in β -cell development and function (13–15). We therefore investigated the WNT-signaling pathway association using summary-level data from an additional 2,625 case subjects and 2,641 control subjects from the DGI (2)

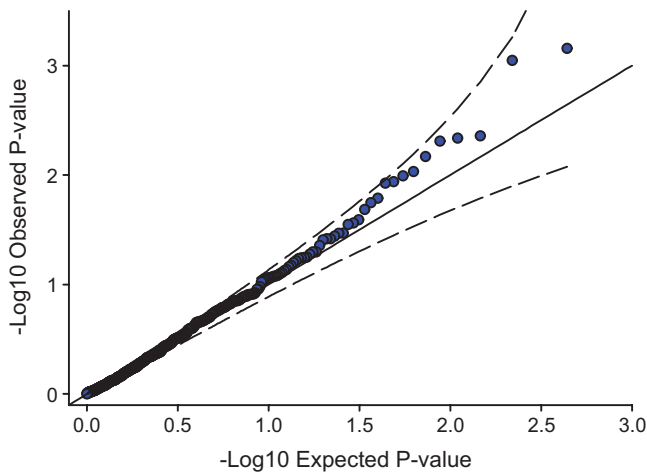


FIG. 1. Dashed lines denote 95% concentration intervals.

and the FUSION (3) studies. We used the type 2 diabetes association statistics available for each of the 126 WNT-signaling gene SNPs. A QQ plot of the meta-analyzed DGI and FUSION results is shown in Fig. 2. There is modest deviation away from the null distribution, but this is within the 95% concentration intervals. Details of the WNT-signaling SNPs reaching $P < 0.01$ are given in Table 2 and for all 126 SNPs in Supplementary Table 1 (found in an online-only appendix at <http://diabetes.diabetesjournals.org/cgi/content/full/db08-1378/DC1>). Expanded results are also available for the next best 12 pathways in Supplementary Table 2.

We next assessed in which pathways the genes in the 18 confirmed type 2 diabetes loci occurred (Supplementary Table 3). Based on regions defined by recombination hotspots, these 18 loci contain 23 genes, although it is important to note that a gene outside of a recombination hotspot may be the gene on which an associated SNP acts. Seven pathways contained more than one gene from the 23 implicated in the 18 loci but none more than three. The highest-ranked of these was the metalloproteinase activity pathway, but this pathway did not reach significance ($P = 0.19$). We also designed three artificial positive control pathways to assess what proportion of real loci need to be present in a pathway for it to be detected. The first

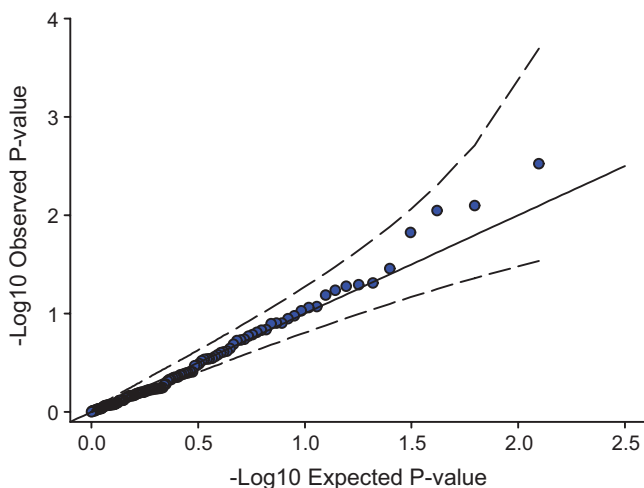


FIG. 2. P values based on meta-analyzed DGI and FUSION data. Dashed lines denote 95% concentration intervals. Plot excludes association signal for *TCF7L2*.

contained the most strongly associated SNP from each of 17 known type 2 diabetes loci (6) (excluding *SLC30A8* not captured on the Affymetrix chip); the second and third contained the same SNPs, diluted with 17 and 34 SNPs with median GWA χ^2 association statistics. Analysis of these 17, 34, and 51 gene pathways, together with the 439 database-defined pathways, ranked each positive control pathway at the top, with nominal P values < 0.0001 . The 17 and 34 gene pathways reached significance after multiple testing (false discovery rate statistics of 0.006 and 0.02, respectively), but the pathway consisting of 51 genes did not (false discovery rate 0.09).

DISCUSSION

The main implication of our results is that type 2 diabetes genes fall in multiple pathways. This is consistent with the observation that, of the 23 genes implicated from the current 18 confirmed loci, a maximum of three fall in the same pathway. There are two main limitations to our study. First, we are limited by the accuracy and completeness of publicly available pathway databases. While defined pathways were used from three separate data sources, there is clearly an incomplete knowledge of biological and disease pathways. Second, results from our artificially created positive control pathways indicate that pathways need to have a large proportion of genes associated before they reach experiment-wise significance (for example, a pathway containing variants representing 17 confirmed loci and 34 nonassociated loci does not reach $P < 0.05$ after correcting for the number of pathways tested).

A further limitation is that the best approach to generate an association statistic for a set of genes is unclear. Large linkage disequilibrium (LD) blocks around strongly associated genes can hugely inflate the significance of an individual pathway and can introduce the problem of nonindependence across genes and pathways. Although this problem was overcome in the GSEA algorithm by selecting a single SNP to represent a gene, this in itself is limiting. A true disease-associated gene may have multiple functional variants. By representing each gene locus with a single SNP, we are therefore missing out on the potential effect of multiple association signals for a given gene. A better gene statistic would evaluate all the variation across a locus, rather than just the SNP with the best association statistic, while accounting for any LD.

Despite our conclusion that diabetes genes are likely to reside in multiple pathways, it is still interesting to note that the second highest-ranked pathway was the WNT-signaling pathway, after the exclusion of *TCF7L2*. Multiple studies have suggested that tight regulation of WNT signaling is required for normal pancreatic development during embryonic growth (13–15). Several WNT-signaling genes, with SNPs associated at nominal $P < 0.01$, are worth noting. These include the *CCND2* gene, a cyclin upregulated in islet compared with whole pancreas (rs11833537 $P = 0.003$), and the *SMAD3* (rs17178347 $P = 0.0006$) and *PRICKLE1* (rs1796390 $P = 0.001$) genes, which are both expressed in the fetal pancreas (data from <http://www.cbil.upenn.edu/epcondb42>).

It was previously shown by Mootha et al. (12), using expression data, that the oxidative phosphorylation pathway was associated with type 2 diabetes. In our analysis, using GWA data, the oxidative phosphorylation pathway ranked 431 of the 439 pathways ($P = 0.98$). The most likely

TABLE 2
Top WNT-SNPs reaching $P < 0.01$ in three study meta-analyses

	SNP	Chromosome	Position	Risk/non-risk allele	Risk frequency	Three study OR	Three study P
<i>TCF7L2</i>	rs4506565	10	114746031	T/A	0.39	1.34 (1.26–1.42)	4.48E-19
<i>SMAD3</i>	rs7178347	15	65108440	T/C	0.35	1.11 (1.05–1.18)	0.0006
<i>PRICKLE1</i>	rs1796390	12	41164871	T/C	0.28	1.11 (1.04–1.19)	0.001
<i>FZD10</i>	rs3741571	12	129161657	T/C	0.06	1.22 (1.08–1.38)	0.002
<i>DKK1</i>	rs1194742	10	53883556	G/A	0.05	1.25 (1.09–1.44)	0.002
<i>SOX17</i>	rs2656272	8	55424453	A/G	0.4	1.1 (1.04–1.17)	0.002
<i>FBXW11</i>	rs33830	5	171080775	A/T	0.36	1.1 (1.03–1.17)	0.003
<i>CSNK2A2</i>	rs2550380	16	56813871	C/T	0.26	1.11 (1.04–1.19)	0.003
<i>CCND2</i>	rs11833537	12	4150368	C/T	0.41	1.09 (1.03–1.16)	0.003
<i>BTRC</i>	rs4436485	10	103235351	G/C	0.63	1.09 (1.03–1.16)	0.004
<i>PPARD</i>	rs9470015	6	35477062	A/G	0.19	1.12 (1.04–1.22)	0.004
<i>SFRP2</i>	rs11732581	4	155077843	C/T	0.96	1.28 (1.09–1.51)	0.004
<i>WNT2B</i>	rs2273368	1	112775813	C/T	0.82	1.1 (1.03–1.18)	0.005
<i>DKK2</i>	rs10028834	4	108211197	T/G	0.08	1.16 (1.04–1.3)	0.007
<i>CSNK1A1L</i>	rs9576222	13	36687910	C/T	0.24	1.09 (1.03–1.16)	0.007
<i>NFAT5</i>	rs16959025	16	68253090	T/G	0.84	1.11 (1.03–1.2)	0.008
<i>CAMK2D</i>	rs6850980	4	115067898	G/A	0.34	1.08 (1.02–1.15)	0.009

Allele frequency is based on WTCCC case samples. Three study P values are from the imputed meta-analysis of WTCCC, DGI, and FUSION cohorts, produced by the DIAGRAM consortium. SNP and gene positions are based on NCBI Build 35. Alleles are based on the forward strand.

reason for this difference is that changes in oxidative phosphorylation gene expression are caused by the disease, rather than the gene expression changes causing type 2 diabetes.

In conclusion, using type 2 diabetes GWA data and a pathway-based approach, we found that the WNT-signaling pathway ranked highest; however, SNPs from this pathway did not show convincing association in additional studies. The failure to find a highly associated individual pathway could be explained by the genetic architecture of type 2 diabetes, the power of the analysis, or the strength and completeness of the algorithms and pathway datasets used. Pathway-based approaches may be more successful for some complex traits than others, depending on the nature of the underlying disease physiology.

ACKNOWLEDGMENTS

Funding for the project was provided by the Wellcome Trust under award 076113. M.N.W. is a Vandervell Foundation Research Fellow.

No potential conflicts of interest relevant to this article were reported.

This study makes use of data generated by the Wellcome Trust Case Control Consortium. A full list of the investigators who contributed to the generation of the data are available from www.wtccc.org.uk.

REFERENCES

- Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007;447:661–678
- Prokopenko I, Langenberg C, Florez JC, Saxena R, Soranzo N, Thorleifsson G, Loos RJ, Manning AK, Jackson AU, Aulchenko Y, Potter SC, Erdos MR, Sanna S, Hottenga JJ, Wheeler E, Kaakinen M, Lyssenko V, Chen WM, Ahmadi K, Beckmann JS, Bergman RN, Bochud M, Bonnycastle LL, Buchanan TA, Cao A, Cervino A, Coin L, Collins FS, Crisponi L, de Geus EJ, Dehghan A, Deloukas P, Doney AS, Elliott P, Freimer N, Gateva V, Herder C, Hofman A, Hughes TE, Hunt S, Illig T, Inouye M, Isomaa B, Johnson T, Kong A, Krestyaninova M, Kuusisto J, Laakso M, Lim N, Lindblad L, Lindgren CM, McCann OT, Mohlke KL, Morris AD, Naitza S, Orrù M, Palmer CN, Pouta A, Randall J, Rathmann W, Saramies J, Scheet P, Scott LJ, Scuteri A, Sharp S, Sijbrands E, Smit JH, Song K, Steinthorsdottir V, Stringham HM, Tuomi T, Tuomilehto J, Uitterlinden AG, Voight BF, Waterworth D, Wichmann HE, Willemsen G, Witteman JC, Yuan X, Zhao

JH, Zeggini E, Schlessinger D, Sandhu M, Boomsma DI, Uda M, Spector TD, Penninx BW, Altshuler D, Vollenweider P, Jarvelin MR, Lakatta E, Waeber G, Fox CS, Peltonen L, Groop LC, Mooser V, Cupples LA, Thorsteinsdottir U, Boehnke M, Barroso I, Van Duijn C, Dupuis J, Watanabe RM, Stefansson K, McCarthy MI, Wareham NJ, Meigs JB, Abecasis GR. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 2007;316:1331–1336

- 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* 2007;316:1341–1345
- 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* 2007;445:881–885
- Steinthorsdottir V, Thorleifsson G, Reynisdottir I, Benediktsson R, Jonsdottir T, Walters GB, Styrkarsdottir U, Gretarsdottir S, Emilsson V, Ghosh S, Baker A, Snorrardottir 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, Thorsteinsdottir U, Kong A, Stefansson K. A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat Genet* 2007;39:770–775
- Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, de Bakker PI, Abecasis GR, Almgren P, Andersen G, Ardlie K, Boström KB, Bergman RN, Bonnycastle LL, Borch-Johnsen K, Burt NP, Chen H, Chines PS, Daly MJ, Deodhar P, Ding CJ, Doney AS, Duren WL, Elliott KS, Erdos MR, Frayling TM, Freathy RM, Gianniny L, Grallert H, Grarup N, Groves CJ, Guiducci C, Hansen T, Herder C, Hitman GA, Hughes TE, Isomaa B, Jackson AU, Jorgensen T, Kong A, Kubalanza K, Kuruvilla FG, Kuusisto J, Langenberg C, Lango H, Lauritzen T, Li Y, Lindgren CM, Lyssenko V, Marville AF, Meisinger C, Midthjell K, Mohlke KL, Morken MA, Morris AD, Narisu N, Nilsson P, Owen KR, Palmer CN, Payne F, Perry JR, Pettersen E, Platou C, Prokopenko I, Qi L, Qin L, Rayner NW, Rees M, Roix JJ, Sandbaek A, Shields B, Sjogren M, Steinthorsdottir V, Stringham HM, Swift AJ, Thorleifsson G, Thorsteinsdottir U, Timpson NJ, Tuomi T, Tuomilehto J, Walker M, Watanabe RM, Weedon MN, Willer CJ; Wellcome Trust Case Control Consortium, Illig T, Hveem K, Hu FB, Laakso M, Stefansson K, Pedersen O, Wareham NJ, Barroso I, Hattersley AT, Collins FS, Groop L, McCarthy MI, Boehnke M, Altshuler D. Meta-analysis of genome-wide association data

- and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet* 2008;40:638–645
7. 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 AS; 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* 2007;316:1336–1341
 8. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology: the Gene Ontology Consortium. *Nat Genet* 2000;25:25–29
 9. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acid Res* 2000;28:27–30
 10. Wang K, Li M, Bucan M. Pathway-based approaches for analysis of genomewide association studies. *Am J Hum Genet* 2007;81:1278–1283
 11. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102:15545–15550
 12. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstrale M, Laurila E, Houtis N, Daly MJ, Patterson N, Mesirov JP, Golub TR, Tamayo P, Spiegelman B, Lander ES, Hirschhorn JN, Altshuler D, Groop LC. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* 2003;34:267–273
 13. Etheridge SL, Spencer GJ, Heath DJ, Genever PG. Expression profiling and functional analysis of wnt signaling mechanisms in mesenchymal stem cells. *Stem Cells* 2004;22:849–860
 14. Papadopoulou S, Edlund H. Attenuated Wnt signaling perturbs pancreatic growth but not pancreatic function. *Diabetes* 2005;54:2844–2851
 15. Ross SE, Hemati N, Longo KA, Bennett CN, Lucas PC, Erickson RL, MacDougald OA. Inhibition of adipogenesis by Wnt signaling. *Science* 2000;289:950–953