Insights into the genetic basis of type 2 diabetes

Norihiro Kato*

ABSTRACT

Type 2 diabetes is one of the most common complex diseases, of which considerable efforts have been made to unravel the pathophysiological mechanisms. Recently, large-scale genome-wide association (GWA) studies have successfully identified genetic loci robustly associated with type 2 diabetes by searching susceptibility variants across the entire genome in an unbiased, hypothesis-free manner. The number of loci has climbed from just three in 2006 to approximately 70 today. For the common type 2 diabetes-associated variants, three features have been noted. First, genetic impacts of individual variants are generally modest; mostly, allelic odds ratios range between 1.06 and 1.20. Second, most of the loci identified to date are not in or near obvious candidate genes, but some are often located in the intergenic regions. Third, although the number of loci is limited, there might be some population specificity in type 2 diabetes association. Although we can estimate a single or a few target genes for individual loci detected in GWA studies by referring to the data for experiments *in vitro*, biological function remains largely unknown for a substantial part of such target genes. Nevertheless, new biology is arising from GWA study discoveries; for example, genes implicated in β-cell dysfunction are over-represented within type 2 diabetes-associated regions. Toward translational advances, we have just begun to face new challenges – elucidation of multifaceted (i.e., molecular, cellular and physiological) mechanistic insights into disease biology by considering interaction with the environment. The present review summarizes recent advances in the genetics of type 2 diabetes, together with its realistic potential. (J Diabetes Invest, doi: 10.1111/jdi.12067, 2013)

KEY WORDS: Genetics, Plasma glucose, Type 2 diabetes

INTRODUCTION

Genetic, environmental and demographic factors, and their interaction, determine an individual's risk for type 2 diabetes; its heritability has been estimated as approximately 25%[1]. Despite considerable concerted efforts over the past 15 years, it is only in the past 5 years that substantial progress has been made in identifying genetic variants robustly associated with type 2 diabetes, largely as a result of technological advances[2]. In particular, the advent of genome-wide association (GWA) studies involving several thousands of samples has facilitated this progress. This approach is to search for susceptibility variants across the entire genome in an unbiased, hypothesis-free manner.

The alleles or mutations responsible for rare monogenic forms of diabetes, including maturity onset diabetes of the young (MODY), were relatively easily identified through family-based linkage analyses[3]. These discoveries then led to molecular diagnostics of the diseases with demonstrable prognostic and therapeutic relevance. Although similar approaches have been applied to common forms of type 2 diabetes, the multifactorial nature has rendered the identification of genetic variants an enormous challenge. In consideration of its low penetrance, it was proposed that association analyses in large unrelated sample sets should be more powerful in susceptibility gene discovery for type 2 diabetes than family-based linkage approaches[4].

The association signal can be detected only if one examines the causal variant itself or a nearby marker with which it is tightly correlated; therefore, researchers were obliged to direct their attention to particular candidate gene variants of interest until the advent of GWA studies. Among a number of candidate genes thus interrogated, common coding variants in *PPARG*[5] and *KCNJ11/ABCC8*[6] were shown to be associated with type 2 diabetes. In 2006, without prior knowledge of biology, *TCF7L2* was first discovered to be a susceptibility gene after systematic association analysis across a 10.5-Mb region of previously reported linkage[7]. Subsequently, in 2007, the first wave of GWA studies identified six novel loci in populations of European descent[8]. Successive rounds of GWA studies and meta-analyses have brought the number of confirmed common variants associated with type 2 diabetes to approximately 70[9–26], and have also discovered >40 common variants influencing normal physiological variation in continuous glycemic measures (e.g., fasting glucose and insulin)[12,27–32] to date.

Department of Gene Diagnostics and Therapeutics, Research Institute, National Center for Global Health and Medicine, Tokyo, Japan
*Corresponding author. Norihiro Kato Tel.: +81-3-3202-7181 Fax: +81-3-3202-7364
E-mail address: nokato@ri.ncgm.go.jp
Received 21 January 2013; revised 25 January 2013; accepted 28 January 2013

In the present article, the evidence in favor of a genetic basis for type 2 diabetes, focusing specifically on the DNA sequence variants that have been implicated in risk predisposition and the assumed clinical implications of genomics, are reviewed.

DISCOVERY OF SUSCEPTIBILITY GENE VARIANTS FOR TYPE 2 DIABETES

Largely through GWA studies, the number of loci robustly implicated in type 2 diabetes risk; that is, those that have attained a genome-wide significance level ($P < 5 \times 10^{-8}$) and also have been repeatedly validated in independent samples, has climbed from just three – *PPARG*, *KCNJ11/ABCC8* and *TCF7L2* – in 2006 to approximately 70 today (Table 1).

GWA studies are based on the principle of linkage disequilibrium (LD) at the population level. LD is the phenomenon in which alleles of two different loci (or genes) occur together more often than would be predicted by chance, indicating that the two alleles are physically close on the DNA strand. LD is created by evolutionary forces, such as mutation, drift and selection, and is broken down by recombination. A set of single nucleotide polymorphisms (SNPs) and mutations in strong LD tend to be inherited together by forming haplotypes. On carrying out GWA studies, we normally assay not all SNPs, but a subset of SNPs that can be chosen by considering the LD structure in a particular chromosomal region. If the pattern and strength of LD between the SNPs and mutations in the target region are similar, and the causal variants are commonly present among different populations (or different ethnic groups), the association in question is detectable at the SNP markers across the populations.

For the common type 2 diabetes-associated variants, three features have to be noted. First, genetic impacts of individual variants or loci are generally modest; that is, allelic odds ratios (ORs) for type 2 diabetes are mostly in the range between 1.06 and 1.20, apart from several loci including *TCF7L2* (Figure 1). This reflects the necessity of a large sample size in meta-analysis to expose variants of smaller effect and more extreme risk allele frequency. Collectively, the most strongly associated variants at individual loci are estimated to explain approximately 10% of familial aggregation of type 2 diabetes[15]. Second, most of the variants identified to date are not in or near obvious candidate genes, but some are often located in the intergenic regions. This leads to the difficulty in estimating causal transcript; that is, the transcript responsible for mediating the effect of the associated variants, according to the location of association signals as discussed later. Third, although the number of variants is limited, there might be some population specificity in type 2 diabetes association. This has to be carefully interpreted by considering several possibilities; for example, the lack of power as a result of insufficient sample size and cross-population differences in LD structure[34]. Although the majority of common variants have a consistent effect on the risk of type 2 diabetes across multiple ethnic groups[35], some variants appear to exert more pronounced genetic effects in specific ethnic groups; for example, the association at *KLF14* is prominent in Europeans[15,36], but not in East[20] and South[19] Asians.

Notably, it has been reported that most of the risk alleles for type 2 diabetes loci share a consistent pattern of decreasing frequencies along human migration from sub-Saharan Africa to East Asia[37]. Such differential frequencies are hypothesized to be caused by the promotion of energy storage and usage appropriate to environments and inconsistent energy intake.

Along with the GWA meta-analyses in individual populations, 'transethnic' meta-analysis is currently being carried out, and will allow for a better chance to show novel susceptibility loci and pathophysiological pathways of type 2 diabetes, and might also facilitate the fine mapping of common causal variants by utilizing ethnic differences in LD structure[38].

OVERLAP OF ASSOCIATION BETWEEN TYPE 2 DIABETES AND GLYCEMIC MEASURES

GWA studies have also identified a number of genetic variants influencing glycemic measures (Tables S1–S3). When we focused on genome-wide significant ($P < 5 \times 10^{-8}$) association signals, we found that they partially overlapped between type 2 diabetes[9–26] and the glycemic measure traits – fasting glucose and insulin/homeostatic model assessment (HOMA)-B (a parameter reflecting β-cell function) and glycated hemoglobin (HbA_{1c}) levels in the non-diabetic population[12,27–32] (Figure 2). As has been pointed out[39], common variants associated with fasting plasma glucose levels do not necessarily influence the risk of type 2 diabetes and, by contrast, those associated with type 2 diabetes do not necessarily influence normal variation in fasting plasma glucose levels (48%, 11 of 23 loci overlapped), suggesting that a different set of genes influence physiological and pathophysiological variation in glucose homeostasis. Furthermore, from the viewpoint of disease mechanism and classification, of interest is the fact that there is some disparity in the list of associated loci between type 2 diabetes and fasting insulin/HOMA-B (45%, 9 of 20 loci overlapped), and between type 2 diabetes and HbA_{1c} levels (38%, 5 of 13 loci overlapped). In the latter case, more than half of the detected loci likely influence HbA_{1c} levels through a non-glycemic pathway, erythrocyte biology (e.g., iron homeostasis)[31].

TRANSITION FROM ASSOCIATION SIGNAL TO CAUSAL MECHANISM

Most loci associated with type 2 diabetes map to regulatory or intergenic regions of the genome, and in many cases the causal transcript remains undetermined. Surprisingly few of the genome-wide association signals have mapped near strong biological candidates. Nevertheless, at some loci, it is inferred, based on a combination of supportive data; for example, coding variants (in particular, non-synonymous SNPs), nearby biological candidates and *cis* expression quantitative trait loci (*cis*-eQTLs), which regulate expression levels of messenger ribonucleic acid. Here, eQTLs that map to the approximate location of their

Table 1 | List of susceptibility loci for type 2 diabetes with significant evidence for association ($P < 5E-8$)

Mapped gene(s)*	Reported gene(s)*	Lead SNP	Region	Pos (GRCh37)	Risk allele	RAF in controls	*P*-value	Reported study†	OR [95% CI]	First-reported ethnic group‡
NOTCH2/ADAM30	NOTCH2	rs10923931	1p12	120517959	T	0.11	4.E-08	Zeggini *et al.*[9]	1.13 [1.08-1.17]	European descent
RPL31P13/PROX1	PROX1	rs340874	1q32.3	214159256	C	0.54	7.E-10	Dupuis *et al.*[12]	1.07 [1.05-1.09]	European descent
GCKR	GCKR	rs780094	2p23.3	27741237	C	0.60	1.E-09	Dupuis *et al.*[12]	1.06 [1.04-1.08]	European descent
THADA	THADA	rs7578597	2p21	43732823	T	0.90	1.E-09	Zeggini *et al.*[9]	1.15 [1.10-1.20]	European descent
EIF3FP3/BCL11A	BCL11A	rs243021	2p16.1	6058481	A	0.48	3.E-15	Voight *et al.*[15]	1.08 [1.06-1.10]	European descent
TMEM163	TMEM163	rs998451	2q21.3	135429288	G	0.86	6.E-12	Tabassum *et al.*[25]	1.56 [1.38-1.77]	South Asian
RND3/FABP5P10	RND3	rs7560163	2q23.3	151637936	C	0.86	7.E-09	Palmer *et al.*[21]	1.33 [1.19-1.49]	African American
RBMS1	RBMS1/ITGB6	rs7593730	2q24.2	161171454	C	0.78	4.E-08	Qi *et al.*[16]	1.11 [1.08-1.16]	European descent
GRB14/COBLL1	GRB14	rs3923113	2q24.3	165501849	A	0.74	1.E-08	Kooner *et al.*[19]	1.09 [1.06-1.13]	South Asian
KIAA1486/IRS1	IRS1	rs7578326	2q36.3	227020653	A	0.65	5.E-20	Voight *et al.*[15]	1.11 [1.08-1.13]	European descent
TIMP4/GSTM5P1§	PPARG	rs13081389	3p25.2	12289800	A	0.96	2.E-07	Voight *et al.*[15]	1.24 [1.15-1.35]	European descent
PSMD6/PRICKLE2	PSMD6	rs831571	3p14.1	64048297	C	0.61	8.E-11	Cho *et al.*[20]	1.09 [1.06–1.12]	East Asian
ADAMTS9/MAGI1	ADAMTS9	rs4607103	3p14.1	64711904	C	0.76	1.E-08	Zeggini *et al.*[9]	1.09 [1.06-1.12]	European descent
ADCY5	ADCY5	rs11708067	3q21.1	123065778	A	0.78	1.E-20	Dupuis *et al.*[12]	1.12 [1.09-1.15]	European descent
IGF2BP2	IGF2BP2	rs4402960	3q27.2	185511687	T	0.31	3.E-09	Perry *et al.*[22]	1.15 [1.10-1.21]	European descent
ST6GAL1	ST6GAL1	rs16861329	3q27.3	186666461	G	0.75	3.E-08	Kooner *et al.*[19]	1.09 [1.06-1.12]	South Asian
WFS1	WFS1	rs1801214	4p16.1	6303022	T	0.73	3.E-08	Voight *et al.*[15]	1.13 [1.08-1.18]	European descent
MAEA	MAEA	rs6815464	4p16.3	1309901	C	0.58	2.E-20	Cho *et al.*[20]	1.13 [1.10–1.16]	East Asian
ANKRD55	ANKRD55	rs459193	5q11.2	55806751	G	0.70	6.E-09	Morris *et al.*[24]	1.08 [1.05–1.11]	European descent
SNORA47/PDE8B	ZBED3	rs4457053	5q13.3	76424949	G	0.26	3.E-12	Voight *et al.*[15]	1.08 [1.06-1.11]	European descent
CDKAL1	CDKAL1	rs7766070	6p22.3	20686573	A	0.27	6.E-11	Perry *et al.*[22]	1.21 [1.14-1.28]	European descent
ZFAND3	ZFAND3	rs9470794	6p21.2	38106844	C	0.27	2.E-10	Cho *et al.*[20]	1.12 [1.08–1.16]	East Asian
KCNK16/KCNK17	KCNK16	rs1535500	6p21.2	39284050	T	0.42	2.E-08	Cho *et al.*[20]	1.08 [1.05–1.11]	East Asian
C6orf57	C6orf57	rs1048886	6q13	71289189	G	0.18	3.E-08	Sim *et al.*[18]	1.54 [1.32-1.80]	Asian Indian
EEF1A1P26/TMEM195	DGKB	rs2191349	7p21.2	15064309	T	0.56	1.E-08	Dupuis *et al.*[12]	1.06 [1.04-1.08]	European descent
JAZF1	JAZF1	rs849134	7p15.1	28196222	A	0.53	3.E-09	Voight *et al.*[15]	1.13 [1.09-1.18]	European descent
GCK/YKT6	GCK	rs4607517	7p13	44235668	A	0.22	5.E-08	Dupuis *et al.*[12]	1.07 [1.05-1.10]	European descent
ZNF800/GCC1	GCC1/PAX4	rs6467136	7q32.1	127164958	G	0.79	5.E-11	Cho *et al.*[20]	1.11 [1.07–1.14]	East Asian
KLF14/FLJ43663	KLF14	rs972283	7q32.3	130466854	G	0.55	2.E-10	Voight *et al.*[15]	1.07 [1.05-1.10]	European descent
ANK1	ANK1	rs516946	8p11.1	41519248	C	0.76	3.E-10	Morris *et al.*[24]	1.09 [1.06–1.12]	European descent
TP53INP1	TP53INP1	rs896854	8q22.1	95960511	T	0.44	1.E-09	Voight *et al.*[15]	1.06 [1.04-1.09]	European descent
SLC30A8	SLC30A8	rs3802177	8q24.11	118185025	G	0.76	1.E-08	Voight *et al.*[15]	1.15 [1.10-1.21]	European descent
UBA52P6/DMRTA1	CDKN2A/2B	rs10965250	9p21.3	22133284	G	0.81	1.E-10	Voight *et al.*[15]	1.20 [1.13-1.27]	European descent
PTPRD	PTPRD	rs17584499	9p24.1	8879118	T	0.06	9.E-10	Tsai *et al.*[13]	1.57 [1.36-1.82]	Han Chinese
GLIS3	GLIS3	rs7041847	9p24.2	4287466	A	0.41	2.E-14	Cho *et al.*[20]	1.10 [1.07–1.13]	East Asian
KRT18P24/CHCHD9	TLE4	rs13292136	9q21.31	81952128	C	0.93	3.E-08	Voight *et al.*[15]	1.11 [1.07-1.15]	European descent
TLE1	TLE1	rs2796441	9q21.32	84308948	G	0.57	5.E-09	Morris *et al.*[24]	1.07 [1.05–1.10]	European descent
CDC123/CAMK1D	CDC123	rs12779790	10p13	12328010	G	0.18	1.E-10	Zeggini *et al.*[9]	1.11 [1.07-1.14]	European descent
VPS26A	VPS26A	rs1802295	10q22.1	70931474	A	0.26	4.E-08	Kooner *et al.*[19]	1.08 [1.05-1.12]	South Asian
ZMIZ1	ZMIZ1	rs12571751	10q22.3	80942631	A	0.52	1.E-10	Morris *et al.*[24]	1.08 [1.05–1.10]	European descent
HHEX/EXOC6	HHEX	rs5015480	10q23.33	94465559	C	0.57	2.E-09	Perry *et al.*[22]	1.18 [1.11-1.23]	European descent

Table 1 (*Continued*)

Mapped gene(s)*	Reported gene(s)*	Lead SNP	Region	Pos (GRCh37)	Risk allele	RAF in controls	*P*-value	Reported study†	OR [95% CI]	First-reported ethnic group‡
TCF7L2	TCF7L2	rs7903146	10q25.2	114758349	T	0.29	2.E-40	Perry *et al.*[22]	1.58 [1.47-1.68]	European descent
GRK5	GRK5	rs10886471	10q26.11	121149403	C	0.78	7.E-09	Li *et al.*[26]	1.12 [1.08-1.16]	Chinese
KCNQ1¶	KCNQ1	rs231362	11p15.5	2691471	G	0.52	3.E-13	Voight *et al.*[15]	1.08 [1.06-1.10]	European descent
KCNQ1	KCNQ1 (OT1)	rs2237895	11p15.4	2857194	C	0.33	1.E-09	Tsai *et al.*[13]	1.29 [1.19-1.40]	Japanese/Han Chinese
KCNJ11/ABCC8	KCNJ11	rs5219	11p15.1	17408630	C	0.40	7.E-11	Zeggini *et al.*[33]	1.14 [1.10-1.19]	European descent
ARAP1	ARAP1	rs1552224	11q13.4	72433098	A	0.87	1.E-22	Voight *et al.*[15]	1.14 [1.11-1.17]	European descent
RPS3AP42/MTNR1B	MTNR1B	rs1387153	11q14.3	92673828	T	0.28	8.E-15	Voight *et al.*[15]	1.09 [1.06-1.11]	European descent
KLHDC5	KLHDC5	rs10842994	12p11.22	27965150	C	0.80	6.E-10	Morris *et al.*[24]	1.10 [1.06–1.13]	European descent
RPSAP52	RPSAP52	rs1531343	12q14.3	66174894	C	0.12	4.E-09	Voight *et al.*[15]	1.10 [1.07-1.14]	European descent
TSPAN8/LGR5	TSPAN8	rs7961581	12q21.1	71663102	C	0.27	1.E-09	Zeggini *et al.*[9]	1.09 [1.06-1.12]	European descent
OASL	OASL	rs7957197	12q24.31	121460686	T	0.85	2.E-08	Voight *et al.*[15]	1.07 [1.05-1.10]	European descent
NDFIP2/SPRY2	SPRY2	rs1359790	13q31.1	80717156	G	0.71	6.E-09	Shu *et al.*[14]	1.15 [1.10-1.20]	Chinese
RASGRP1	RASGRP1	rs7403531	15q14	38822905	T	0.35	4.E-09	Li *et al.*[26]	1.10 [1.06-1.13]	Chinese
C2CD4A/C2CD4B	C2CD4A/4B	rs7172432	15q22.2	62396389	A	0.58	9.E-14	Yamauchi *et al.*[17]	1.11 [1.08-1.14]	Japanese
HMG20A	HMG20A	rs7178572	15q24.3	77747190	G	0.52	7.E-11	Kooner *et al.*[19]	1.09 [1.06–1.12]	South Asian
ZFAND6/FAH	ZFAND6	rs11634397	15q25.1	80432222	G	0.64	2.E-09	Voight *et al.*[15]	1.06 [1.04-1.08]	European descent
AP3S2	AP3S2	rs2028299	15q26.1	90374257	C	0.31	2.E-11	Kooner *et al.*[19]	1.10 [1.07-1.13]	South Asian
PRC1	PRC1	rs8042680	15q26.1	91521337	A	0.22	2.E-10	Voight *et al.*[15]	1.07 [1.05-1.09]	European descent
FTO	FTO	rs11642841	16q12.2	53845487	A	0.45	3.E-08	Voight *et al.*[15]	1.13 [1.08-1.18]	European descent
BCAR1	CTRB1/B2	rs7202877	16q23.1	75247245	T	0.89	4.E-08	Morris *et al.*[24]	1.12 [1.07–1.16]	European descent
SRR	SRR	rs391300	17p13.3	2216258	G	0.62	3.E-09	Tsai *et al.*[13]	1.28 [1.18-1.39]	Han Chinese
HNF1B	HNF1B	rs4430796	17q12	36098040	G	0.28	2.E-11	Li *et al.*[26]	1.19 [1.13-1.25]	European descent
LAMA1	LAMA1	rs8090011	18p11.31	7068462	G	0.38	8.E-09	Perry *et al.*[22]	1.13 [1.09-1.18]	European descent
MC4R	MC4R	rs12970134	18q22	57884750	A	0.27	1.E-08	Morris *et al.*[24]	1.08 [1.05–1.11]	European descent
CILP2	CILP2	rs10401969	19p13.11	19407718	C	0.08	7.E-09	Morris *et al.*[24]	1.13 [1.09–1.18]	European descent
PEPD	PEPD	rs3786897	19q13.11	33893008	A	0.56	1.E-08	Cho *et al.*[20]	1.10 [1.07–1.14]	East Asian
FITM2/R3HDML	FITM2	rs6017317	20q13.12	42946966	G	0.48	1.E-11	Cho *et al.*[20]	1.09 [1.07–1.12]	East Asian
HNF4A	HNF4A	rs4812829	20q13.12	42989267	A	0.29	3.E-10	Kooner *et al.*[19]	1.09 [1.06-1.12]	South Asian
KRT18P48/DUSP9	DUSP9	rs5945326	Xq28	152899922	A	0.79	3.E-10	Voight *et al.*[15]	1.27 [1.18-1.37]	European descent

Mapped and reported genes are arbitrarily denoted according to those shown in A Catalog of Published Genome-Wide Association Studies (http://www.genome.gov/gwastudies/index.cfm?pageid=26525384#searchForm). †When >1 studies have reported genome-wide significant association at the relevant loci, we select one for each locus according to: (i) an ethnic group where the association was first reported; and (ii) the largest study in a given ethnic group. ‡At KCNQ1, although the association was first reported in Japanese (ref. 10,11), it is not included in A Catalog of Published Genome-Wide Association Studies and we show a Chinese study (ref. 13) alternatively for reference. §On chromosome 3p25.2, although it did not attain a genome-wide significance level in each study, the reproducible association has been shown for a candidate gene, PPARG, rs13081389, which is in linkage disequilibrium (r^2 = 0.536) with rs1801282 (P12A, PPARG) in HapMap CEU. RAF,risk allele frequency; SNP, single nucleotide polymorphism. ¶At KCNQ1, significant association was also reported in European-descent populations (ref. 15), which is not in linkage disequilibrium with the one-first reported in Japanese (ref. 10,11).

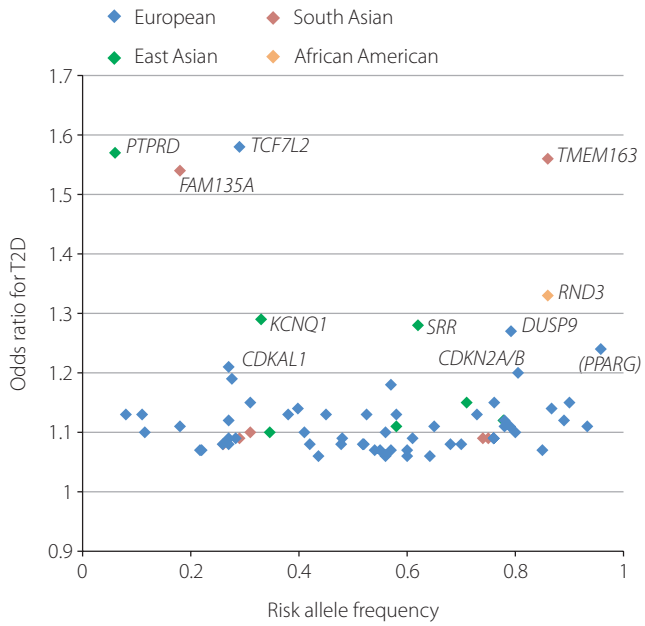

Figure 1 | Risk allele frequencies and effect sizes of known susceptibility loci for type 2 diabetes (T2D), which have shown significant ($P \leq 5 \times 10^{-8}$) association. Gene names are attached to the loci with odds ratio (OR) ≥1.2; they are not necessarily proven to be causal, but represent candidate transcripts on the basis of location and biological plausibility. Although a large part of associated loci were originally identified in populations of European descent, some were exclusively found or first reported in non-European populations, which are differentially colored in the figure.

gene-of-origin are referred to as *cis*-eQTLs. Although eQTLs can be used to identify the downstream targets that are likely to be affected by associations detected in GWA studies, they still rely on genotyping methods, and therefore point to regions of LD rather than to individual SNPs. Accordingly, independent methods for identifying SNPs that overlap regulatory elements, such as transcription factor binding sites, are required. High-throughput functional assays (e.g., ChIP-seq[40]) can experimentally detect functional chromosomal regions, such as transcription factor binding sites, and the presence of SNPs in these regions can lead to differences in transcription factor binding between individuals[41]. Recently, the relevant experimental datasets have been generated and released to the public by the Encyclopedia of DNA Elements (ENCODE) Consortium[40], which will help identify functional SNPs associated with type 2 diabetes and their potential causal transcript.

Although we can estimate a single or a few target genes for individual loci by referring to the data for experiments *in vitro*; for example, the ENCODE data, biological function remains largely unknown for a substantial part of such target genes. So far, just 17% (12 of 70 loci) have been proven to show type 2 diabetes-related phenotypes in their knock-out mice experiments *in vivo* (Table 2). Besides, three target genes – *GCK*, *HNF1B* and *HNF4A* – overlap with causal genes for MODY, where *GCK*[42] and *HNF4A*[43] knock-out mice show hyperglycemia and glucose intolerance, respectively.

Despite significant enrichment for regulatory (and non-coding) sequence variants in disease-associated regions, there are some cases where substantial statistical and biological evidence can support particular coding sequence variants as causal. For example, the type 2 diabetes association signal on chromosome 2p23 was shown to derive from a common non-synonymous SNP rs1260326, P446L, in *GCKR*, which is one of 17 genes mapped to the 420-kb interval of association in tight LD[44]. In addition to the strong candidacy of *GCKR* in glucose metabolism, functional characterization *in vitro* showed that P446L could explain a mutational mechanism for the reported counter-intuitive association with increased triglycerides and reduced glucose levels on 2p23.

NEW BIOLOGY ARISING FROM GWA STUDY DISCOVERIES

β-Cell Dysfunction

Regarding the pathogenesis of type 2 diabetes, there has been a long-standing debate over the relative roles of insulin secretory defects and insulin resistance. In this context, of interest is the fact that a large part of the type 2 diabetes-risk loci exert their primary effects on disease risk through reduced insulin secretion rather than increased insulin resistance in the general population[15,45]. Genes implicated in cell-cycle regulation are overrepresented within type 2 diabetes-associated regions; this is consistent with the notion that control of β-cell mass is a key component of disease risk[15].

However, when we look at genetic loci associated with proinsulin levels, there are divergent directions of association between type 2 diabetes risk and proinsulin levels[46]. Similar to the relationship between type 2 diabetes and fasting plasma glucose[39], the loci are partially overlapped between the traits. Among the loci associated with proinsulin, three loci – *TCF7L2*, *C2CD4A* and *SLC30A8* – were significantly associated with type 2 diabetes in a manner consistent with established epidemiological relationships; that is, higher proinsulin levels are associated with impaired β-cell function, insulin resistance and risk of type 2 diabetes[47,48]. In contrast, one locus – *ARAP1* – showed trait association in a counterinstinctive direction. Thus, both disproportionate elevations and reductions in proinsulin can indicate β-cell dysfunction at individual loci[46].

Efforts to show that the genes mapping close to susceptibility loci are enriched for particular pathways or processes have not been particularly rewarding so far, apart from a few instances (e.g., cell-cycle regulation)[15]. This indicates the possibility that type 2 diabetes is highly heterogeneous, and/or existing biological knowledge is as yet insufficient to capture key fundamental aspects of its pathophysiology through database search. Although it is challenging to establish the biological mechanism at each associated locus, a combination of experimental and bioinformatic approaches will help understand the broad

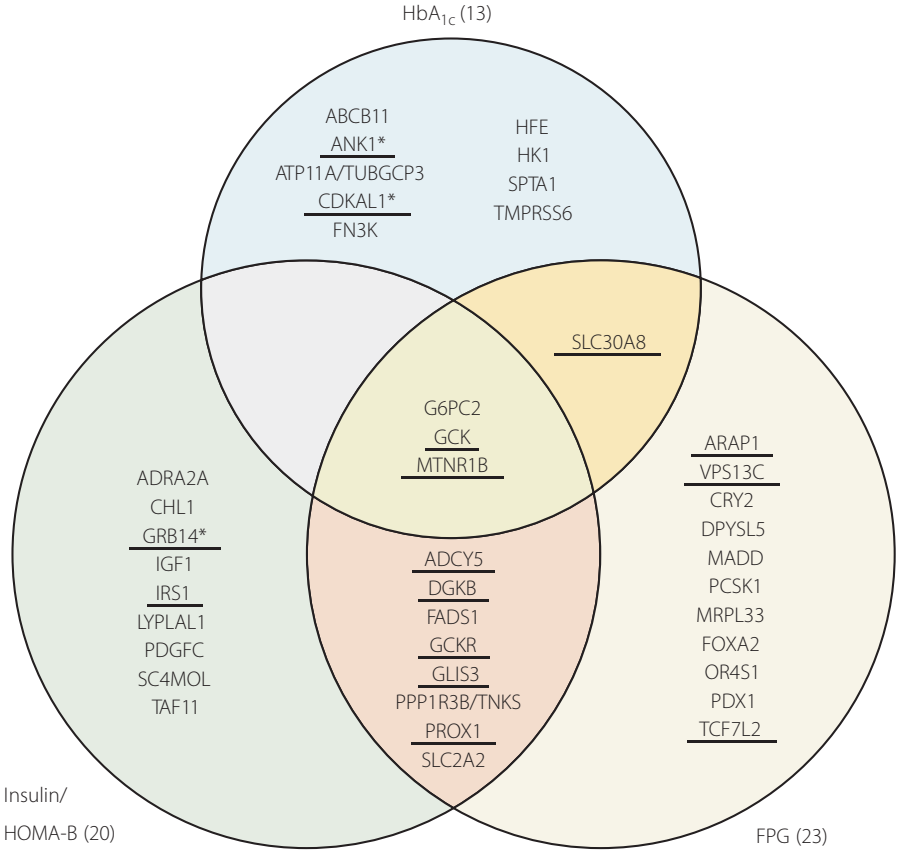

Figure 2 | A schematic representation of intertrait difference (or overlapping) for 41 diabetes-related trait associated loci that have been reported in meta-analyses of genome-wide association studies[12,27–32]. The traits include fasting plasma glucose (FPG), insulin and its related-traits (homeostasis model assessment of β-cell function (HOMA-B) and HOMA of insulin resistance), and glycated hemoglobin (HbA_{1c}). Here, an associated locus is assumed to overlap between the traits when $P \leq 5 \times 10^{-8}$ was concordantly attained. Underlined are the loci that have shown significant ($P \leq 5 \times 10^{-8}$) association with type 2 diabetes; at three loci with asterisks – *ANK1*, *CDKAL1* and *GRB14* – variants associated with individual traits are not in linkage disequilibrium ($r^2 < 0.3$).

processes of disease pathogenesis by integrating a number of loci identified in the unbiased genome-wide approach.

Epigenetics

The evidence for familial aggregation of type 2 diabetes comes from a number of epidemiological studies; parental type 2 diabetes has been reported to give rise to an approximately three-fold increase in disease risk in the offspring[49]. The familial aggregation might reflect epigenetic mechanisms, such as the fetal origins hypothesis[50], in addition to genetic influences and shared family environment. As an approach to addressing this issue, Kong *et al.*[51] examined the impact of parental origin on disease associations in previous GWA studies and identified parental-origin-specific associations with type 2 diabetes at variants located in the known imprinted region on chromosome 11p15. Here, the allele that confers risk when paternally inherited (odds ration [OR] = 1.41, $P = 4.3 \times 10^{-9}$) is protective when maternally transmitted (OR = 0.87, $P = 0.02$) and also correlated with decreased methylation of CTCF-binding site at 11p15[51].

A growing body of data has established that the molecular basis of metabolic programming involves DNA methylation and histone modifications[52]. To date, relatively few studies have explored the epigenetic component to the development of type 2 diabetes, with most of them focusing on the methylation status of selected C-phosphate-G (CpG) sites in candidate genes. Because of the relatively high cost and procedural complexity of epigenetic analysis, as well as tissue differences in methylation profile, there are few convincing results that support the contribution of epigenetics to disease pathogenesis at present. To make the situation intricate, it has been reported at *FTO*, one of the principal risk loci for obesity and type 2 diabetes, that epigenetic effects might, at least in part, be driven by underlying variation in the DNA sequence[53]. That is, methylation levels at a CpG site in the first intron of *FTO* were correlated with a genotype at nearby disease-associated SNPs. This could be simply regarded as a subset of genetic association signals at which the downstream effects are mediated by genotype-dependent changes in local DNA methylation. However, it remains unclear whether methylation by

Table 2 | Genome-wide association study-identified positional candidate genes for type 2 diabetes, with supportive phenotypes observed in knock-out mice

Gene	MGI ID	Phenotypes observed in knock-out mice	Reference (PMID no.)	Other associated trait identified via GWA study*
GRB14	1355324	Improved glucose tolerance, insulin levels decreased, increased incorporation of glucose into glycogen in the liver and skeletal muscle of males. Both males and females showed a decrease in body size.	Cooney GJ, 2004 *EMBO J* (14749734)	Waist-hip ratio Blood pressure
IRS1	99454	Impaired glucose tolerance, mild insulin and IGF-1 resistance; 50% reductions in body weight at birth and at 4 months-of-age. Homozygotes: lethal.	Araki E, 1994 *Nature* (7526222)	Visceral adipose tissue/ subcutaneous adipose tissue ratio Adiponectin levels
PPARG	97747	Heterozygotes: greater β-cell proliferation, enhanced leptin secretion, and resistance to high-fat diet-induced adipocyte hypertrophy and insulin resistance.	Kubota N, 1999 *Molecular Cell* (10549291)	Plasminogen activator inhibitor type 1 levels
WFS1	1328355	Decreased pancreatic beta cells, impaired glucose tolerance, decreased body weight and abnormal behavior associated with increased sensitivity to stress.	Ishihara H, 2004 *Hum Mol Genet* (15056606); Riggs AC, 2005 *Diabetologia* (16215705)	N/A
SLC30A8	2442682	Reduced islet zinc levels, insulin levels decreased and glucose-stimulated insulin secretion decreased.	Lemaire K, 2009 *Proc Natl Acad Sci USA* (19706465)	Asthma
GLIS3	2444289	Postnatal lethality associated with neonatal diabetes and polycystic kidney disease.	Kang HS, 2009 *Mol Cell Biol* (19273592); Watanabe N, 2009 *FEBS Lett* (19481545)	Type 1 diabetes
FTO	1347093	Body weight decreased, adipose tissue decreased and body fat decreased; metabolism increased, serum lipids increased and serum glucagon increased.	Fischer J, 2009 *Nature* (19234441)	Body mass index Waist circumference Osteoarthritis Menarche
MC4R	99457	Hyperglycemia and weight gain.	Huszar D, 1997 Cell (9019399)	Body mass index Waist circumference Height
HNF4A	109128	Nullizygous embryos: delayed growth and lethality. Conditional deletion in pancreatic beta cells: hyperinsulinemia and impaired glucose tolerance.	Gupta RK, 2005 J Clin Invest (15761495); Pearson ER, 2007 *PLoS Med* (17407387)	C-reactive protein Ulcerative colitis
GCKR	1096345	Reduced glucokinase protein levels and activity in the liver and altered glucose homeostasis.	Farrelly D, 1999 *Proc Natl Acad Sci USA* (10588736)	Total protein/albumin levels Sex hormone-binding globulin levels Phospholipid levels Platelet counts C-reactive protein Crohn's disease Urate levels Chronic kidney disease
GCK	1270854	Mild hyperglycemia in heterozygous mice and extreme hyperglycemia and embryonic to postnatal ethality in homozygous mice.	Bali D, 1995 *J Biol Chem* (7665557)	N/A
CDKAL1	1921765	Conditional deletion in pancreatic beta cells: impaired tRNA Lys modification, reduction of glucose-stimulated proinsulin synthesis. Global deletion: body weight decresed, glucose intolerance manifested after 20 weeks of high-fat diet.	Wei FY, 2011 *J Clin Invest* (21841312); Okamura T, 2012 *PLoS One* (23173044)	Body mass index Birth weight Crohn's disease

*Phenotype traits, with which genome-wide association (GWA) studies identified significant association at the corresponding gene locus are listed, except for lipid and glucose-related traits. IGF-1, insulin-like growth factor 1; MGI, mouse genome informatics.

Table 3 | List of gene variants showing potential pleiotropic effects on type 2 diabetes and other traits

Trait	Nearby gene(s)	Variant	LD coefficient, r^2 (HapMap panel)	Type 2 diabetes-associated SNP*	Effect on the trait†	Reported study
Adiponectin levels	*IRS1*	rs925735	0.648 (CEU)	rs7578326	↓	Dastani Z, 2012 *PLoS Genet*
Adiponectin levels	*PEPD*	rs731839	0.345 (CEU) 0.894 (JPT+CHB)	rs3786897 (East Asians)	↓	Dastani Z, 2012 *PLoS Genet*
Birthweight	*ADCY5*	rs9883204	0.782 (CEU)	rs11708067	↓	Freathy RM, 2010 *Nat Genet*
Birthweight	*CDKAL1*	rs6931514	1.000 (CEU)	rs7766070	↓	Horikoshi M, 2012 *Nat Genet*
Height	*IGF2BP2*	rs720390	0.491 (CEU)	rs4402960	↑	Lango Allen H, 2010 *Nature*
Height	*JAZF1*	rs1635852	1.000 (CEU)	rs849134	↑	Johansson A, 2008 Hum *Mol Genet*
Height	*C2CD4A*	rs7178424	0.422 (CEU) 0.082 (JPT+CHB)	rs7172432 (Japanese)	Unknown	Lango Allen H, 2010 *Nature*
Height	*MC4R*	rs17782313	0.813 (CEU)	rs12970134	↑	Lango Allen H, 2010 *Nature*
Type 1 diabetes	*GLIS3*	rs7020673	0.902 (CEU) 0.705 (JPT+CHB)	rs7041847 (East Asians)	↑	Barrett JC, 2009 *Nat Genet*
Type 1 diabetes	*RASGRP1*	rs8035957	0.733 (CEU) 0.414 (JPT+CHB)	rs7403531 (Chinese)	Unknown	Grant SF, 2008 *Diabetes*
Multiple sclerosis	*HHEX*	rs7923837	0.699 (CEU)	rs5015480	↑	Sawcer S, 2011 *Nature*
Coronary heart disease	*SRR/SMG6*	rs216172	0.552 (CEU) 0.588 (JPT+CHB)	rs391300 (Chinese)	↓	Schunkert H, 2011 *Nat Genet*
Prostate cancer /endometrial cancer	*HNF1B*	rs4430796	1.000 (CEU)	rs4430796	↓	Gudmundsson J, 2007 *Nat Genet*

In the table, we list type 2 diabetes-associated variants, whose proxy single nucleotide polymorphisms (SNPs; $r^2 > 0.4$) also show significant association with other traits. *In some cases, significant type 2 diabetes-association was reported only in East Asian populations, whereas the association with other trait(s) was found in populations of European-descent. †The direction of association was estimated by considering the linkage disequilbrium (LD) information on haplotypes except for two loci (C2CD4A and RASGRP1), where LD was too modest.

itself constitutes the causal link between the *FTO* risk allele and type 2 diabetes[54].

Pleiotropy

GWA studies of type 2 diabetes have provided substantial evidence of pleiotropy; the same variants are associated with multiple traits (Table 3), providing clues to the common biological pathways involved. For example, at *ADCY5* and *CDKAL1*, the birth weight-lowering allele was associated with a greater risk of type 2 diabetes[55]. This is consistent with the fetal insulin hypothesis[56]; that is, common genetic variation influencing insulin secretion or action, both in prenatal development and adult life, could partly explain epidemiological correlations between lower birth weight and type 2 diabetes. Here, of particular note is the fact that the type 2 diabetes risk allele at *CDKAL1* also showed a significant association with lower body mass index (BMI) in adult East Asians[57,58], indicating the possibility of sustained, reduced insulin secretion in adulthood. In this line, despite the clustering of type 2 diabetes and obesity in metabolic syndrome, the directions of association appear to be divergent at susceptibility loci. A positive correlation has been shown between type 2 diabetes risk and higher BMI for a few obesity-associated loci; for example, *FTO* and *MC4R*, whereas an inverse correlation between the traits has been found for several type 2 diabetes-associated loci including *CDKAL1*[59]. This reveals further complexity in biological pathways that influence metabolic impairments.

Several studies have examined the association of type 2 diabetes variants with cardiovascular outcome, assuming some causal association between the diseases[60,61]. A potentially shared association has been noted at a genomic region near the *SRR* locus, where the minor allele C of rs216172 is positively associated with coronary heart disease (CHD) risk (OR = 1.07) in the population of European-descent[62], and the minor allele A of rs391300 is inversely associated with type 2 diabetes (OR = 0.78) in the Chinese population[13]; two SNPs near *SRR* are in LD ($r^2 = 0.552$ in the HapMap population of European ancestry and 0.588 in the HapMap population of East Asian ancestry). At another region on chromosome 9p21 near *CDKN2A/B*, a significant association has been identified for CHD[62] and type 2 diabetes[15,19,20], as well as several other diseases. The 9p21 region is under intense research[63], as independent functional variants are likely to exist within this region and could be associated with individual diseases. Apart from these two loci, the results for associations between the individual diabetes-predisposing genetic variants and CHD risk appear to be inconsistent. A few recent studies, however, have shown that, when tested in aggregate using a genetic risk score, the overall genetic predisposition to type 2 diabetes is associated with an increased risk of CHD[60,61].

GWA studies have provided evidence for an interrelation between type 2 diabetes and prostate cancer[64]. Observational studies have consistently shown an inverse association between the two diseases, with meta-analysis risk ratios ranging from 0.84 to 0.91[65,66]. In good accordance with this, one shared genomic region at *HNF1B* has been highlighted in GWA scan[67], where the major allele A of rs4430796 is positively associated with prostate cancer (OR = 1.22) and inversely associated with type 2 diabetes risk (OR = 0.91). Although the biological mechanism underlying such paradoxical associations is poorly understood, it is hypothesized that in type 2 diabetes patients, metabolic status might move gradually from hyperinsulinemia to endogenous insulin deficiency, thus blunting oncogenic action of insulin in the prostate[68]. The GWA results have further indicated a direct association between diabetes risk variants other than *HNF1B* and prostate cancer risk plus the lack of significant evidence supporting the potential for a type 2 diabetes phenotype to mediate the genetic effect of *HNF1B*[64].

POTENTIAL OF TRANSLATIONAL ADVANCES

Missing Heritability

It has been argued that susceptibility loci identified through GWA studies explain only a small proportion of heritability (approximately 5–10%) for type 2 diabetes. This discrepancy, termed missing heritability[69], has been attributed to a number of factors including insufficient survey of rare variants and structural variants, inaccuracy of current heritability estimates (e.g., inflation because of shared family environment), and epigenetics. It remains to be defined whether complex traits are truly affected by thousands of variants with small effect, but recent analysis of GWA study data using a computational technique has suggested that many hundreds of common weakly-associated variants might be sufficient to account for the majority of heritability – approximately 50% of overall trait variance for type 2 diabetes[70], in accordance with the assumption of 'hidden' heritability[71]. Along this, given the strong interplay of genetics, epigenetics and environment, partitioning individual propensity to develop type 2 diabetes is not feasible.

Personalized Medicine

Successful applications of personalized medicine in the clinical management of diabetes patients are restricted to the rare, monogenic forms of disease. For example, it is known that MODY patients with *HNF1A* mutations respond particularly well to sulphonylurea treatment[72]. Similar efforts have been made for common forms of type 2 diabetes, in two principal areas of personalized medicine – molecular prediction (or diagnosis) and personalized therapy. To provide improved predictive power over conventional risk factors, genetic testing must be sensitive and specific in discriminating subjects who will develop the disease on follow up from those who will not[73]. In this line, it has been recognized that genetic variants so far identified do not substantially improve the discriminative accuracy of disease prediction based on conventional risk factors[74,75]. Even genetic models incorporating thousands of additional putative common variants are likely to offer limited improvement[73,76]. Although some studies have shown that molecular prediction is slightly more effective in certain patient groups, such as the young[74,77], the discriminative accuracy still falls short of the clinical utility at the individual level. At the group level, in contrast, risk stratification is achievable to some extent by using a genetic risk score (GRS); this is an integrated summary of genetic risk from all the different variants in the genome that GWA studies have identified as predisposing to the disease. The GRS thus calculated has the capacity to highlight patient groups at the top end of the risk distribution[78,79]. A higher GRS was shown to be associated with indices of diminished β-cell function and incidence of diabetes during follow up, gaining predictive ability in comparison with clinical characteristics alone[78]. Furthermore, lifestyle interventions appear to be effective even among individuals at highest genetic risk[78,80]. Therefore, it is worth testing whether targeting such high-risk groups for an earlier preventative intervention strategy is beneficial.

From the viewpoint of individual therapeutic utility, pharamacogenetic studies in common forms of type 2 diabetes have not yet achieved remarkable progress, apart from a few successes in the candidate gene approach; for example, positive associations between variation in sulfonylurea response and genotype at the *ABCC8/KCNJ11* and *TCF7L2* loci[81,82]. It is assumed that individual loci affecting antidiabetes drug response exert modest effects, and hence large-scale pharamacogenetic GWA studies are required to identify novel susceptibility gene variants.

Also, in terms of clinical management, genetics of diabetic microvascular complications – retinopathy, nephropathy and neuropathy – is an issue of great interest. Although a number of suggestive loci have been nominated through candidate gene approach or GWA study[83,84], none have attained genome-wide significant association with the disease, partly because of the lack of statistical power. Given the substantial genetic heterogeneity, future large-scale consortium-based studies are warranted.

PERSPECTIVE

Remarkable progress has been made in the genetics of type 2 diabetes in the past 5 years, principally through GWA studies. This proceeds with the rapid technological advances in 'the era of big data', which will further enable the sequencing of entire genomes in large samples at affordable costs. We expect that a larger list of associated loci can be discovered in the next few years, thanks to unprecedented global collaboration involving different ethnic groups. Under such circumstances, we have just begun to face new challenges – elucidation of multifaceted (i.e., molecular, cellular and physiological) mechanistic insights into disease biology by considering interaction with environment[85] – before clinical translation.

ACKNOWLEDGEMENTS

This work was supported by a grant from the Program for Promotion of Fundamental Studies in Health Sciences of the National Institute of Biomedical Innovation Organization and a Grant of National Center for Global Health and Medicine. The authors declare no conflict of interest.

REFERENCES

1. Poulsen P, Kyvik KO, Vaag A, *et al.* Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance–a population-based twin study. *Diabetologia* 1999; 42: 139–145.
2. Visscher PM, Brown MA, McCarthy MI, *et al.* Five years of GWAS discovery. *Am J Hum Genet* 2012; 90: 7–24.
3. Owen K, Hattersley AT. Maturity-onset diabetes of the young: from clinical description to molecular genetic characterization. *Best Pract Res Clin Endocrinol Metab* 2001; 15: 309–323.
4. Merikangas KR, Risch N. Genomic priorities and public health. *Science* 2003; 302: 599–601.
5. Altshuler D, Hirschhorn JN, Klannemark M, *et al.* The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat Genet* 2000; 26: 76–80.
6. Gloyn AL, Weedon MN, Owen KR, *et al.* 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* 2003; 52: 568–572.
7. Grant SF, Thorleifsson G, Reynisdottir I, *et al.* Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat Genet* 2006; 38: 320–323.
8. Travers ME, McCarthy MI. Type 2 diabetes and obesity: genomics and the clinic. *Hum Genet* 2011; 130: 41–58.
9. Zeggini E, Scott LJ, Saxena R, *et al.* 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.
10. Yasuda K, Miyake K, Horikawa Y, *et al.* Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. *Nat Genet* 2008; 40: 1092–1097.
11. Unoki H, Takahashi A, Kawaguchi T, *et al.* SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations. *Nat Genet* 2008; 40: 1098–1102.
12. Dupuis J, Langenberg C, Prokopenko I, *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* 2010; 42: 105–116.
13. Tsai FJ, Yang CF, Chen CC, *et al.* A genome-wide association study identifies susceptibility variants for type 2 diabetes in Han Chinese. *PLoS Genet* 2010; 6: e1000847.
14. Shu XO, Long J, Cai Q, *et al.* Identification of new genetic risk variants for type 2 diabetes. *PLoS Genet* 2010; 6: e1001127.
15. Voight BF, Scott LJ, Steinthorsdottir V, *et al.* Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet* 2010; 42: 579–589.
16. Qi L, Cornelis MC, Kraft P, *et al.* Genetic variants at 2q24 are associated with susceptibility to type 2 diabetes. *Hum Mol Genet* 2010; 19: 2706–2715.
17. Yamauchi T, Hara K, Maeda S, *et al.* A genome-wide association study in the Japanese population identifies susceptibility loci for type 2 diabetes at UBE2E2 and C2CD4A-C2CD4B. *Nat Genet* 2010; 42: 864–868.
18. Sim X, Ong RT, Suo C, *et al.* Transferability of type 2 diabetes implicated loci in multi-ethnic cohorts from Southeast Asia. *PLoS Genet* 2011; 7: e1001363.
19. Kooner JS, Saleheen D, Sim X, *et al.* Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. *Nat Genet* 2011; 43: 984–989.
20. Cho YS, Chen CH, Hu C, *et al.* Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. *Nat Genet* 2011; 44: 67–72.
21. Palmer ND, McDonough CW, Hicks PJ, *et al.* A genome-wide association search for type 2 diabetes genes in African Americans. *PLoS ONE* 2012; 7: e29202.
22. Perry JR, Voight BF, Yengo L, *et al.* Stratifying type 2 diabetes cases by BMI identifies genetic risk variants in LAMA1 and enrichment for risk variants in lean compared to obese cases. *PLoS Genet* 2012; 8: e1002741.
23. Imamura M, Maeda S, Yamauchi T, *et al.* A single-nucleotide polymorphism in ANK1 is associated with susceptibility to type 2 diabetes in Japanese populations. *Hum Mol Genet* 2012; 21: 3042–3049.
24. Morris AP, Voight BF, Teslovich TM, *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet* 2012; 44: 981–990.
25. Tabassum R, Chauhan G, Dwivedi OP, *et al.* Genome-wide Association Study for type 2 diabetes in Indians identifies a new susceptibility locus at 2q21. *Diabetes* 2013; 62: 977–986.
26. Li H, Gan W, Lu L, *et al.* A genome-wide association study identifies GRK5 and RASGRP1 as type 2 diabetes loci in Chinese Hans. *Diabetes* 2013; 62: 291–298.
27. Manning AK, Hivert MF, Scott RA, *et al.* A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat Genet* 2012; 44: 659–669.
28. Chen G, Bentley A, Adeyemo A, *et al.* Genome-wide association study identifies novel loci association with fasting insulin and insulin resistance in African Americans. *Hum Mol Genet* 2012; 21: 4530–4536.
29. Paré G, Chasman DI, Parker AN, *et al.* Novel association of HK1 with glycated hemoglobin in a non-diabetic population: a genome-wide evaluation of 14,618

participants in the Women's Genome Health Study. *PLoS Genet* 2008; 4: e1000312.
30. Franklin CS, Aulchenko YS, Huffman JE, *et al.* The TCF7L2 diabetes risk variant is associated with HbA (C) levels: a genome-wide association meta-analysis. *Ann Hum Genet* 2010; 74: 471–478.
31. Soranzo N, Sanna S, Wheeler E, *et al.* Common variants at 10 genomic loci influence hemoglobin A (C) levels via glycemic and nonglycemic pathways. *Diabetes* 2010; 59: 3229–3239.
32. Ryu J, Lee C. Association of glycosylated hemoglobin with the gene encoding CDKAL1 in the Korean Association Resource (KARE) study. *Hum Mutat* 2012; 33: 655–659.
33. Zeggini E, Weedon MN, Lindgren CM, *et al.* Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 2007; 316: 1336–1341.
34. Kato N. Ethnic differences in genetic predisposition to hypertension. *Hypertens Res* 2012; 35: 574–581.
35. Waters KM, Stram DO, Hassanein MT, *et al.* Consistent association of type 2 diabetes risk variants found in europeans in diverse racial and ethnic groups. *PLoS Genet* 2010; 6: e1001078.
36. Small KS, Hedman AK, Grundberg E, *et al.* Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes. *Nat Genet* 2011; 43: 561–564.
37. Chen R, Corona E, Sikora M, *et al.* Type 2 diabetes risk alleles demonstrate extreme directional differentiation among human populations, compared to other diseases. *PLoS Genet* 2012; 8: e1002621.
38. Kato N. Ethnic diversity in type 2 diabetes genetics between East Asians and Europeans. *J Diabetes Invest* 2012; 3: 349–351.
39. De Silva NM, Frayling TM. Novel biological insights emerging from genetic studies of type 2 diabetes and related metabolic traits. *Curr Opin Lipidol* 2010; 21: 44–50.
40. ENCODE Project Consortium, Dunham I, Kundaje A, *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012; 489: 57–74.
41. Kasowski M, Grubert F, Heffelfinger C, *et al.* Variation in transcription factor binding among humans. *Science* 2010; 328: 232–235.
42. Inoue M, Sakuraba Y, Motegi H, *et al.* A series of maturity onset diabetes of the young, type 2 (MODY2) mouse models generated by a large-scale ENU mutagenesis program. *Hum Mol Genet* 2004; 13: 1147–1157.
43. Miura A, Yamagata K, Kakei M, *et al.* Hepatocyte nuclear factor-4alpha is essential for glucose-stimulated insulin secretion by pancreatic beta-cells. *J Biol Chem* 2006; 281: 5246–5257.
44. Orho-Melander M, Melander O, Guiducci C, *et al.* Common missense variant in the glucokinase regulatory protein gene is associated with increased plasma triglyceride and C-reactive protein but lower fasting glucose concentrations. *Diabetes* 2008; 57: 3112–3121.
45. Perry JR, Frayling TM. New gene variants alter type 2 diabetes risk predominantly through reduced beta-cell function. *Curr Opin Clin Nutr Metab Care* 2008; 11: 371–377.
46. Strawbridge RJ, Dupuis J, Prokopenko I, *et al.* Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes. *Diabetes* 2011; 60: 2624–2634.
47. Røder ME, Porte D Jr, Schwartz RS, *et al.* Disproportionately elevated proinsulin levels reflect the degree of impaired B cell secretory capacity in patients with noninsulin-dependent diabetes mellitus. *J Clin Endocrinol Metab* 1998; 83: 604–608.
48. Wareham NJ, Byrne CD, Williams R, *et al.* Fasting proinsulin concentrations predict the development of type 2 diabetes. *Diabetes Care* 1999; 22: 262–270.
49. Meigs JB, Cupples LA, Wilson PW. Parental transmission of type 2 diabetes: the Framingham Offspring Study. *Diabetes* 2000; 49: 2201–2207.
50. Hales CN, Barker DJ. Type 2 (non-insulin-dependent) diabetes mellitus: the thrifty phenotype hypothesis. *Diabetologia* 1992; 35: 595–601.
51. Kong A, Steinthorsdottir V, Masson G, *et al.* Parental origin of sequence variants associated with complex diseases. *Nature* 2009; 462: 868–874.
52. Handel AE, Ebers GC, Ramagopalan SV. Epigenetics: molecular mechanisms and implications for disease. *Trends Mol Med* 2010; 16: 7–16.
53. Bell CG, Finer S, Lindgren CM, *et al.* Integrated genetic and epigenetic analysis identifies haplotype-specific methylation in the FTO type 2 diabetes and obesity susceptibility locus. *PLoS ONE* 2010; 5: e14040.
54. Toperoff G, Aran D, Kark JD, *et al.* Genome-wide survey reveals predisposing diabetes type 2-related DNA methylation variations in human peripheral blood. *Hum Mol Genet* 2012; 21: 371–383.
55. Horikoshi M, Yaghootkar H, Mook-Kanamori DO, *et al.* New loci associated with birth weight identify genetic links between intrauterine growth and adult height and metabolism. *Nat Genet* 2012; 45: 76–82.
56. Hattersley AT, Tooke JE. The fetal insulin hypothesis: an alternative explanation of the association of low birthweight with diabetes and vascular disease. *Lancet* 1999; 353: 1789–1792.
57. Wen W, Cho YS, Zheng W, *et al.* Meta-analysis identifies common variants associated with body mass index in east Asians. *Nat Genet* 2012; 44: 307–311.
58. Okada Y, Kubo M, Ohmiya H, *et al.* Common variants at CDKAL1 and KLF9 are associated with body mass index in east Asian populations. *Nat Genet* 2012; 44: 302–306.

59. Okamura T, Yanobu-Takanashi R, Takeuchi F, *et al.* Deletion of CDKAL1 affects high-fat diet-induced fat accumulation and glucose-stimulated insulin secretion in mice, indicating relevance to diabetes. *PLoS One* 2012; 7: e49055.
60. Pfister R, Barnes D, Luben RN, *et al.* Individual and cumulative effect of type 2 diabetes genetic susceptibility variants on risk of coronary heart disease. *Diabetologia* 2011; 54: 2283–2287.
61. Qi Q, Meigs JB, Rexrode KM, *et al.* Diabetes genetic predisposition score and cardiovascular complications among patients with type 2 diabetes. *Diabetes Care* 2013; 36: 737–739.
62. Schunkert H, König IR, Kathiresan S, *et al.* Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet* 2011; 43: 333–338.
63. Harismendy O, Notani D, Song X, *et al.* 9p21 DNA variants associated with coronary artery disease impair interferon-γ signalling response. *Nature* 2011; 470: 264–268.
64. Machiela MJ, Lindström S, Allen NE, *et al.* Association of type 2 diabetes susceptibility variants with advanced prostate cancer risk in the breast and prostate cancer cohort consortium. *Am J Epidemiol* 2012; 176: 1121–1129.
65. Bonovas S, Filioussi K, Tsantes A. Diabetes mellitus and risk of prostate cancer: a meta-analysis. *Diabetologia* 2004; 47: 1071–1078.
66. Kasper JS, Giovannucci E. A meta-analysis of diabetes mellitus and the risk of prostate cancer. *Cancer Epidemiol Biomarkers Prev* 2006; 15: 2056–2062.
67. Gudmundsson J, Sulem P, Steinthorsdottir V, *et al.* Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat Genet* 2007; 39: 977–983.
68. Giovannucci E, Michaud D. The role of obesity and related metabolic disturbances in cancers of the colon, prostate, and pancreas. *Gastroenterology* 2007; 132: 2208–2225.
69. Manolio TA, Collins FS, Cox NJ, *et al.* Finding the missing heritability of complex diseases. *Nature* 2009; 461: 747–753.
70. Stahl EA, Wegmann D, Trynka G, *et al.* Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nat Genet* 2012; 44: 483–489.
71. Gibson G. Hints of hidden heritability in GWAS. *Nat Genet* 2010; 42: 558–560.
72. Pearson ER, Starkey BJ, Powell RJ, *et al.* Genetic cause of hyperglycaemia and response to treatment in diabetes. *Lancet* 2003; 362: 1275–1281.
73. Jostins L, Barrett JC. Genetic risk prediction in complex disease. *Hum Mol Genet* 2011; 20: R182–R188.
74. Meigs JB, Shrader P, Sullivan LM, *et al.* Genotype score in addition to common risk factors for prediction of type 2 diabetes. *N Engl J Med* 2008; 359: 2208–2219.
75. Raynor LA, Pankow JS, Duncan BB, *et al.* Novel risk factors and the prediction of type 2 diabetes in the Atherosclerosis Risk in Communities (ARIC) study. *Diabetes Care* 2013; 36: 70–76.
76. Evans DM, Visscher PM, Wray NR. Harnessing the information contained within genome-wide association studies to improve individual prediction of complex disease risk. *Hum Mol Genet* 2009; 18: 3525–3531.
77. de Miguel-Yanes JM, Shrader P, Pencina MJ, *et al.* Genetic risk reclassification for type 2 diabetes by age below or above 50 years using 40 type 2 diabetes risk single nucleotide polymorphisms. *Diabetes Care* 2011; 34: 121–125.
78. Hivert MF, Jablonski KA, Perreault L, *et al.* Updated genetic score based on 34 confirmed type 2 diabetes Loci is associated with diabetes incidence and regression to normoglycemia in the diabetes prevention program. *Diabetes* 2011; 60: 1340–1348.
79. Villegas R, Delahanty R, Gao YT, *et al.* Joint effect of genetic and lifestyle risk factors on type 2 diabetes risk among Chinese men and women. *PLoS ONE* 2012; 7: e49464.
80. Florez JC, Jablonski KA, Bayley N, *et al.* TCF7L2 polymorphisms and progression to diabetes in the Diabetes Prevention Program. *N Engl J Med* 2006; 355: 241–250.
81. Sesti G, Laratta E, Cardellini M, *et al.* The E23K variant of KCNJ11 encoding the pancreatic beta-cell adenosine 5′-triphosphate-sensitive potassium channel subunit Kir6.2 is associated with an increased risk of secondary failure to sulfonylurea in patients with type 2 diabetes. *J Clin Endocrinol Metab* 2006; 91: 2334–2339.
82. Jablonski KA, McAteer JB, de Bakker PI, *et al.* Common variants in 40 genes assessed for diabetes incidence and response to metformin and lifestyle intervention in the diabetes prevention program. *Diabetes* 2010; 59: 2672–2681.
83. Grassi MA, Tikhomirov A, Ramalingam S, *et al.* Genome-wide meta-analysis for severe diabetic retinopathy. *Hum Mol Genet* 2011; 20: 2472–2481.
84. Palmer ND, Freedman BI. Insights into the genetic architecture of diabetic nephropathy. *Curr Diab Rep* 2012; 12: 423–431.
85. Manning AK, LaValley M, Liu CT, *et al.* Meta-analysis of gene-environment interaction: joint estimation of SNP and SNP × environment regression coefficients. *Genet Epidemiol* 2011; 35: 11–18.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Table S1 | List of loci regulating fasting plasma glucose level with suggestive or significant evidence for association ($P < 1E\text{-}5$)
Table S2 | List of loci regulating insulin-related traits with suggestive or significant evidence for association ($P < 1E\text{-}5$)
Table S3 | List of loci regulating glycated hemoglobin level with suggestive or significant evidence for association ($P < 1E\text{-}5$)