

## ANNALS OF THE NEW YORK ACADEMY OF SCIENCES

Issue: *The Year in Diabetes and Obesity***The genetics of type 2 diabetes: what have we learned from GWAS?**Liana K. Billings<sup>1,2,3</sup> and Jose C. Florez<sup>1,2,3,4</sup>

<sup>1</sup>Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts. <sup>2</sup>Department of Medicine, Harvard Medical School, Boston, Massachusetts. <sup>3</sup>Diabetes Research Center (Diabetes Unit), Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts. <sup>4</sup>Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts

Address for correspondence: Jose C. Florez, M.D., Ph.D., Center for Human Genetic Research, Simches Research Building, CPZN 5.250, Massachusetts General Hospital, 185 Cambridge Street, Boston, MA 02114. jcflorez@partners.org

Type 2 diabetes mellitus has been at the forefront of human diseases and phenotypes studied by new genetic analyses. Thanks to genome-wide association studies, we have made substantial progress in elucidating the genetic basis of type 2 diabetes. This review summarizes the concept, history, and recent discoveries produced by genome-wide association studies for type 2 diabetes and glycemic traits, with a focus on the key notions we have gleaned from these efforts. Genome-wide association findings have illustrated novel pathways, pointed toward fundamental biology, confirmed prior epidemiological observations, drawn attention to the role of  $\beta$ -cell dysfunction in type 2 diabetes, explained  $\sim 10\%$  of disease heritability, tempered our expectations with regard to their use in clinical prediction, and provided possible targets for pharmacotherapy and pharmacogenetic clinical trials. We can apply these lessons to future investigation so as to improve our understanding of the genetic basis of type 2 diabetes.

**Keywords:** genome-wide association studies; type 2 diabetes; genetics; single nucleotide polymorphisms

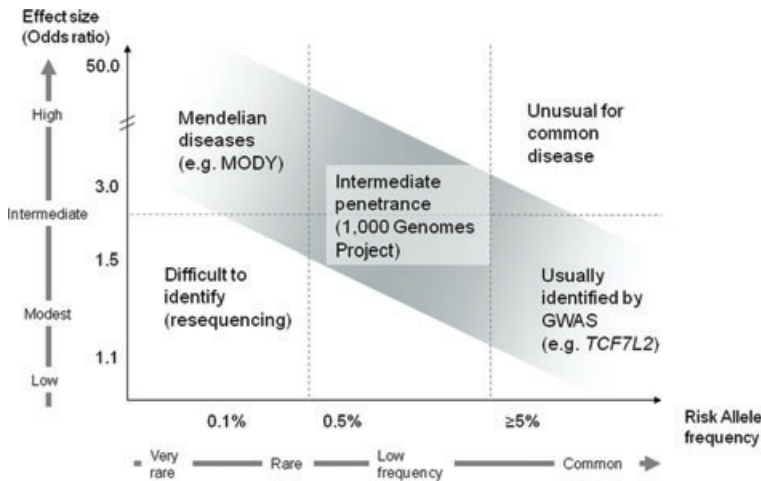
**Introduction**

During the last decade there has been an outpouring of studies providing clues into the genetic architecture underlying complex diseases. Type 2 diabetes mellitus (T2D) has been at the forefront of human diseases and traits studied by new genetic analyses. Prior to genome-wide association studies (GWAS), the primary methods used to establish a link between genotype and phenotype were linkage analysis and candidate gene approaches. Linkage analysis relies on shared DNA segments inherited from common ancestors coupled with phenotypic information. This method was useful in identifying familial genetic variants with large effects, such as those giving rise to maturity-onset diabetes of the young (MODY).<sup>1</sup> When applied to common T2D, Reynisdottir *et al.* identified segments in chromosomes 5 and 10 with suggestive linkage to T2D.<sup>2</sup> The chromosome 10 region harbored *TCF7L2*, one of the T2D-associated genes with strongest effects,

although not sufficiently high to explain the original linkage signal.<sup>3</sup>

Linkage analysis can detect rare genetic loci that strongly influence a disease, but proved limited in unveiling common genetic variants with a more modest impact on complex diseases (Fig. 1). Association methods could detect genetic variants causing smaller effects. Unlike linkage analysis, in their earlier form, association methods could only be applied to candidate genes; they were therefore biased in assuming that a specific locus caused disease based on biological plausibility, and limited in that they depended on prior knowledge with a consequent disregard for intergenic regions. From this work, *PPARG* and *KCNJ11* emerged as two candidate genes, both of which encode targets for anti-diabetes medications and harbor missense variants associated with T2D.<sup>4–6</sup>

Thanks to the completion of the Human Genome and International HapMap Projects (see below), the novel approach of searching for genetic



**Figure 1.** Frequency of genetic variation and disease susceptibility. A portion of “missing heritability” may be explained by low-frequency variants with intermediate penetrance. The low-risk-allele frequencies make the variants undetectable by current GWAS arrays. Resequencing technology and new-generation arrays may help identify these low-frequency variants; larger sample sizes may be needed to detect significant signals. The effort can be further strengthened by applying prior biological and epidemiological data to select associations. Figure adapted from Refs. 113 and 114.

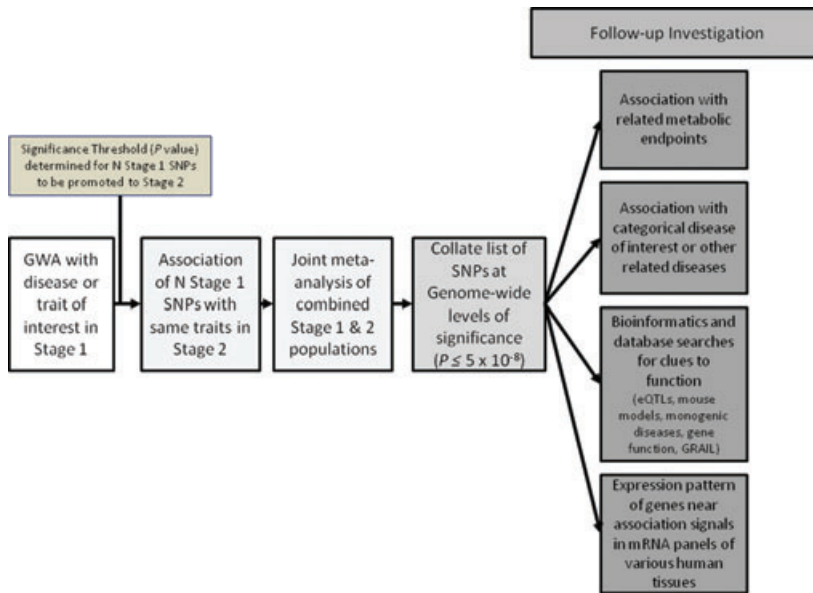
associations in a “genome-wide” fashion came to fruition (Fig. 2). Thus, scientists embarked in GWAS, which allowed them to discover multiple gene variants with individually small effects. Once a specific polymorphism is associated with a disease, it is usually annotated by naming the gene in closest proximity to it. However, this does not necessarily mean that the variant in question is the molecular defect responsible for the phenotype, nor does it implicate the nearest gene; it simply flags a genomic region that harbors the causal variant, which may itself be acting at a certain distance, for instance, by modulating expression of a far-away gene. Therefore, while association signals are often identified by gene names, only in a few cases has a causal relationship been demonstrated, typically via fine-mapping and functional approaches.

Substantial progress in our knowledge of the genetic basis of T2D has been elucidated by T2D GWAS, but there remains a large portion of unexplained genetic heritability. This review summarizes the concept, history, and recent discoveries produced by GWAS for T2D and glycemic traits. It will emphasize what we have learned from the explosion of data in the last 10 years, and how we can apply our knowledge to future investigation to gain further understanding of the genetic basis of T2D.

## Launching GWAS

The ability to interrogate the entire genome was made possible by two key advances: the Human Genome Project, with a draft sequence in 2001 and near-complete sequence in 2003, and the International HapMap project, with its first phase completed in 2005, and now currently in its third phase.<sup>7–10</sup> Progress in high-throughput and affordable genotyping technology; analytical tools to assist in the data mining, cleaning, and interpretation of large databases; and the assembly of international collaborations combining well-phenotyped cohorts made using these advances possible.

Following the completion of the Human Genome Project, a search for genetic variation that might explain phenotypic diversity and an individual’s risk of disease ensued. Assaying single nucleotide polymorphisms (SNPs) became a mainstream way to study the association of genetic variation and disease. A SNP is a single nucleotide DNA sequence variant in the genome that differs between members of the same species or a pair of chromosomes in an individual. SNPs occur on average every 300 base pairs, have a low rate of recurrent mutation, and are most often binary in nature. Several million SNPs were discovered and deposited in public databases.<sup>11</sup> Initially, the HapMap genotyped 3.9 million SNPs in 270 DNA samples among four different ethnic groups



**Figure 2.** Schematic of a typical GWAS design. The general design of GWAS starts with a stage 1 (discovery) cohort. The top SNPs are promoted to the stage 2 (replication) cohort based on (1) a significance threshold that is usually dictated by pragmatic considerations, (2) whether there is prior knowledge of association between the disease the variant, and (3) whether an association is biologically plausible. The successfully replicated SNPs are meta-analyzed in the combined stage 1 and stage 2 cohorts. The SNPs that reach levels of genome-wide significance ( $P = 5 \times 10^{-8}$ ) are explored further using the functional analysis techniques listed.

and defined the underlying patterns of the inheritance of genetic variation. The inheritance pattern is quantified by linkage disequilibrium (LD), which represents the likelihood that alleles of nearby SNPs will stay together and preserve their linear arrangement on a haplotype during meiosis. This likelihood is dependent on recombination rates, with recombination events more likely to separate alleles that lie further apart. In this manner, two SNPs in strong LD will be inherited together more frequently than two SNPs in weak LD. By knowing this correlation structure, investigators and chip manufacturers only have to query a smaller subset of SNPs, or “tag” SNPs, to design genotyping arrays and conduct association analyses that essentially capture the majority of remaining common genomic variation. Genetic variants that are not directly genotyped can then be imputed from the genotyped “tag” SNPs subset. Imputation presumes the allele of a SNP at a different location inferred by its degree of LD with an allele at a directly genotyped variant.<sup>12</sup>

### Initial GWAS for T2D

GWAS have led to the discovery of 38 SNPs associated with T2D, in addition to nearly two dozen SNPs associated with glycemic traits (Table 1). Notably,

although these SNPs have been associated with the trait of interest beyond reasonable statistical doubt, save a few exceptions, they have not yet led to the identification of the specific causal variant(s).

The first GWAS for T2D was conducted in a French discovery cohort composed of 661 cases of T2D (body mass index [BMI] < 30 kg/m<sup>2</sup>, first-degree family history of T2D) and 614 nondiabetic controls were genotyped on two genotyping platforms. In total, 392,935 SNPs culled from two different genotyping platforms were analyzed for association with T2D. Although two associations were not reproducible in follow-up studies (*LOC387761*, *EXT2*), this study identified novel and reproducible association signals at *SLC30A8* and *HHEX* and validated the well known association at *TCF7L2*.<sup>13</sup> Investigators from the Icelandic company deCODE and their collaborators confirmed the association of loci *SLC30A8* and *HHEX* with T2D and identified an additional signal in *CDKAL1*.<sup>14</sup> On the same day, three collaborating groups, the Wellcome Trust Case Control Consortium (WTCCC), the Finland-United States Investigation of NIDDM Genetics (FUSION) group, and the Diabetes Genetics Initiative (DGI), published their findings replicating *SLC30A8* and *HHEX*, and independently

**Table 1.** Genetic variants associated with T2D or glycemic traits at genome-wide levels of statistical significance

Locus	Marker	Chr	Type of mutation	Allele (Effect/Other)	Effect allele frequency (HapMap-CEU)	Trait	Effect: OR/Beta (95% CI/SE)
2000							
<i>PPARG</i>	rs1801282 <sup>4</sup>	3	Missense: Pro12Ala	C/ G	0.92	T2D	1.14 (1.08–1.20) <sup>15–17</sup>
2003							
<i>KCNJ11/ ABCC8</i>	rs5219/ rs757110 <sup>a,5</sup>	11	Missense: Glu23Lys/Ala1369Ser	T/ C G/ T	0.50 0.40	T2D	1.15 (1.09–1.21) <sup>17</sup>
2006							
<i>TCF7L2</i>	rs7903146 <sup>b,3,35</sup>	10	Intronic	T/ C	0.25	T2D FG	1.37 (1.28–1.47) <sup>18</sup> 0.023 (0.004)
	rs12243326 <sup>b,36</sup>		Intronic	C/ T	0.21	2-hr G	0.07 (0.01)
2007							
<i>IGF2BP2</i>	rs4402960 <sup>15–17</sup>	3	Intronic	T/ G	0.29	T2D	1.17 (1.10–1.25) <sup>18</sup>
<i>CDKAL1</i>	rs7754840 <sup>15–17</sup>	6	Intronic	C/ G	0.31	T2D	1.12 (1.08–1.16)
<i>SLC30A8</i>	rs13266634 <sup>13,35</sup>	8	Missense: Arg325Trp	C/ T	0.75	T2D FG	1.12 (1.07–1.16) <sup>65</sup> 0.027 (0.004)
<i>CDKN2A/ B</i>	rs10811661 <sup>15–17</sup>	9	125 kb upstream	T/ C	0.79	T2D	1.20 (1.14–1.25)
<i>HHEX</i>	rs1111875 <sup>13</sup>	10	7.7 kb downstream	C/ T	0.56	T2D	1.13 (1.08–1.17) <sup>17</sup>
<i>FTO</i>	rs8050136 <sup>15,16</sup>	16	Intronic	A/ C	0.45	T2D	1.15 (1.09–1.22) <sup>18</sup>
<i>HNF1B</i>	rs757210 <sup>110,111</sup>	17	Intronic	A/ G	0.43	T2D	1.12 (1.07–1.18) <sup>65</sup>
2008							
<i>NOTCH2</i>	rs10923931 <sup>18</sup>	1	Intronic	T/ G	0.11	T2D	1.13 (1.08–1.17)
<i>THADA</i>	rs7578597 <sup>18</sup>	2	Missense: Thr1187Ala	T/ C	0.92	T2D	1.15 (1.10–1.20)
<i>ADAMSTS9</i>	rs4607103 <sup>18</sup>	3	38 kb upstream	C/ T	0.81	T2D	1.09 (1.06–1.12)
<i>JAZF1</i>	rs864745 <sup>18</sup>	7	Intronic	T/ C	0.52	T2D	1.10 (1.07–1.13)
<i>CDC123/ CAMK1D</i>	rs12779790 <sup>18</sup>	10	Intergenic region	G/ A	0.23	T2D	1.11 (1.07–1.14)
<i>KCNQ1</i>	rs2237892 <sup>c,24</sup>	11	Intronic	C/ T	0.61 <sup>d</sup>	T2D	1.4 (1.34–1.47)
	rs231362 <sup>c,65</sup>		Intronic	G/ A	0.52	T2D	1.08 (1.06–1.10)
<i>TSPAN8/ LGR5</i>	rs7961581 <sup>18</sup>	12	Intronic	C/ T	0.23	T2D	1.09 (1.06–1.12)
2009							
<i>IRS1</i>	rs2943641 <sup>27</sup>	2	502 kb upstream	C/ T	0.61	T2D	1.19 (1.13–1.25)
2010							
<i>DUSP9</i>	rs5945326 <sup>65</sup>	X	8 kb upstream	G/ A	0.12	T2D	1.27 (1.18–1.37)
<i>PROX1</i>	rs340874 <sup>35</sup>	1	2 kb upstream	C/ T	0.50	FG T2D	0.013 (0.003) 1.07 (1.05–1.09)
<i>BCL11A</i>	rs243021 <sup>65</sup>	2	99 kb downstream	A/ G	0.46	T2D	1.08 (1.06–1.10)
<i>G6PC2</i>	rs560887 <sup>35</sup>	2	Intronic	C/ T	0.70	FG HOMA-B HbA <sub>1c</sub>	0.075 (0.004) –0.042 (0.004) 0.032 (0.004)
<i>GCKR</i>	rs1260326 <sup>e,17</sup>	2	Missense: Leu446Pro	T/ C	0.40	2-hr G	0.10 (0.01)
	rs780094 <sup>e,35</sup>		Intronic	C/ T	0.62	FG HOMA-IR T2D	0.029 (0.003) 0.032 (0.004) 0.035 (0.004) 1.06 (1.04–1.08)
<i>ADCY5</i>	rs2877716 <sup>f,36,81</sup>	3	Intronic	C/ T	0.77	2-hr G	0.07 (0.01)
	rs11708067 <sup>f,35</sup>		Intronic	A/ G	0.78	FG HOMA-B T2D	0.027 (0.003) –0.023 (0.004) 1.12 (1.09–1.15)

*Continued*

**Table 1. Continued**

Locus	Marker	Chr	Type of mutation	Allele (Effect/Other)	Effect allele frequency (HapMap-CEU)	Trait	Effect: OR/Beta (95% CI/SE)
<i>SLC2A2</i>	rs11920090 <sup>35</sup>	3	Intronic	T/ A	0.85	FG	0.02 (0.004)
<i>WFS1</i>	rs1801214 <sup>65</sup>	4	Intron-exon junction	G/ A	0.27	T2D	1.13 (1.07–1.18)
<i>ZBED3</i>	rs4457053 <sup>65</sup>	5	41 kb upstream	G/ A	0.26	T2D	1.08 (1.06–1.11)
<i>DGKB/</i>	rs2191349 <sup>35</sup>	7	Intergenic region	T/ G	0.47	FG	0.03 (0.003)
<i>TMEM195</i>						T2D	1.06 (1.04–1.08)
<i>GCK</i>	rs4607517 <sup>35</sup>	7	36 kb upstream	A/ G	0.20	FG	0.062 (0.004)
						HbA <sub>1C</sub>	0.041 (0.005)
						T2D	1.07 (1.05–1.10)
<i>KLF14</i>	rs972283 <sup>65</sup>	7	47 kb upstream	G/ A	0.55	T2D	1.07 (1.05–1.10)
<i>TP53INP1</i>	rs896854 <sup>65</sup>	8	Intronic	T/ C	0.48	T2D	1.06 (1.04–1.09)
<i>GLIS3</i>	rs7034200 <sup>35</sup>	9	Intronic	A/ C	0.53	FG	0.018 (0.003)
						HOMA-B	–0.020 (0.004)
<i>TLE4<sup>8</sup></i>	rs13292136 <sup>65</sup>	9	234 kb upstream	C/ T	0.93	T2D	1.11 (1.07–1.15)
<i>ADRA2A</i>	rs10885122 <sup>35</sup>	10	210 kb downstream	G/ T	0.90	FG	0.022 (0.004)
<i>CENTD2</i>	rs1552224 <sup>65</sup>	11	5' UTR	A/ C	0.88	T2D	1.14 (1.11–1.17)
<i>CRY2</i>	rs11605924 <sup>35</sup>	11	Intronic	A/ C	0.54	FG	0.015 (0.003)
<i>FADS1</i>	rs174550 <sup>35</sup>	11	Intronic	T/ C	0.63	FG	0.017 (0.003)
						HOMA-B	–0.020 (0.003)
<i>MADD</i>	rs7944584 <sup>35</sup>	11	Intronic	A/ T	0.69	FG	0.021 (0.003)
<i>MTNR1B</i>	rs10830963 <sup>35</sup>	11	Intronic	G/ C	0.30	FG	0.067 (0.003)
						HOMA-B	–0.034 (0.004)
						HbA <sub>1C</sub>	0.024 (0.004)
						T2D	1.09 (1.06–1.12)
<i>HMGA2</i>	rs1531343 <sup>65</sup>	12	43 kb upstream	C/ G	0.10	T2D	1.10 (1.07–1.14)
<i>HNFA1A</i>	rs7957197 <sup>65</sup>	12	20 kb downstream	T/ A	0.85	T2D	1.07 (1.05–1.10)
<i>IGF1</i>	rs35767 <sup>35</sup>	12	1.2 kb upstream	G/ A	0.90	FI	0.01 (0.006)
						HOMA-IR	0.013 (0.006)
<i>C2CD4B</i>	rs11071657 <sup>35</sup>	15	21 kb downstream	A/ G	0.59	FG	0.008 (0.003)
<i>PRC1</i>	rs8042680 <sup>65</sup>	15	Intronic	A/ C	0.22	T2D	1.07 (1.05–1.09)
<i>VPS13C</i>	rs17271305 <sup>36</sup>	15	Intronic	G/ A	0.42	2-hr G	0.07 (0.01)
<i>ZFAND6</i>	rs11634397 <sup>65</sup>	15	1.5 kb downstream	G/ A	0.56	T2D	1.06 (1.04–1.08)
<i>GIPR</i>	rs10423928 <sup>36</sup>	19	Intronic	A/ T	0.18	2-hr G	0.11 (0.01)

Loci are arranged by year of genome-wide significance (GWS). The SNP citations reference the publications where the SNP initially associated with the trait. FG, fasting glucose (mmol/ L); FI, fasting insulin (pmol/ L); 2-hr G, 2-hour glucose (FG adjusted, mmol/L); HbA<sub>1C</sub>, Hemoglobin A<sub>1C</sub> (%). Effect estimates are adapted from the references for each SNP unless otherwise noted.

<sup>a</sup>These two missense SNPs are in strong LD in two adjacent genes. Their protein products form the two components of the hetero-octameric beta-cell K<sub>ATP</sub> channel. Recent functional analysis suggests that the missense mutation at *ABCC8* is likely the causal variant for gliclazide response.

<sup>b</sup>These SNPs are in strong LD in Europeans ( $r^2 = 0.79$ ).

<sup>c</sup>These SNPs are in low LD in Europeans ( $r^2 = 0.01$ ) and likely represent independent association signals.

<sup>d</sup>Allele frequency from HapMap-JPT.

<sup>e</sup>These SNPs are in strong LD in Europeans ( $r^2 = 0.93$ ) but not in Africans ( $r^2 = 0.43$ ).

<sup>f</sup>These SNPs are in strong LD in Europeans ( $r^2 = 0.82$ ).

<sup>g</sup>Formerly annotated as *CHCHD9*.

discovering novel associations at *CDKALI*, *IGF2BP2*, and *CDKN2A/B*.<sup>15–17</sup>

### More power to detect low-effect size common variants with increasing sample size

WTCCC, FUSION, and DGI ultimately combined their data to form the Diabetes Genetics Replication and Metaanalysis (DIAGRAM) consortium, which led to a substantial increase in sample size and, thus, power to detect common genetic variants with low effect size.<sup>18</sup> Table 1 summarizes the findings of each of these cohorts and consortia.

Most T2D genetics cohorts have now coalesced to form DIAGRAM+, which achieved an effective sample size of over 22,000 subjects of European origin. In a recent report, 2,426,886 imputed and genotyped autosomal SNPs, with additional interrogation of the X-chromosome, were examined for association with T2D as a categorical phenotype. Fourteen signals, also shown in Table 1, reached genome-wide significance ( $P < 5 \times 10^{-8}$ ) in association with T2D (this threshold is determined by correction for the estimated  $10^6$  independent tests that capture common genomic variation in non-African populations<sup>19</sup>).

In addition to nine novel loci, this most recent meta-analysis allowed for the confirmation of other loci previously associated with T2D among smaller cohorts including *IRS1*, *MTNR1B*, and *KCNQ1*.<sup>20–25</sup> In independent work, *IRS1* (encoding the insulin receptor substrate-1) had been associated with T2D, development of hyperglycemia, insulin resistance by homeostasis model assessment (HOMA-IR<sup>26</sup>), fasting glucose, and fasting insulin.<sup>27</sup> Several GWAS for fasting glucose as a quantitative trait, described in more detail below, had already identified *MTNR1B* as a locus influencing fasting hyperglycemia, making it a candidate locus for association with T2D.<sup>21–23</sup> A third SNP, rs231362, is located in an intron of *KCNQ1* in chromosome 11, which overlaps the *KCNQ1OT1* transcript, thought to influence expression of *CDKN1C*, which regulates  $\beta$ -cell development;<sup>28</sup> an independent signal in *KCNQ1* had been associated with T2D in Japanese, Korean, Chinese, and European populations.<sup>24,25</sup> Two additional loci were near genes that had been linked with T2D in previous studies, but not at genome-wide significance: *HNF1A* harbors rare mutations that account for MODY, and *BCL11A* had

shown suggestive association not reaching genome-wide significance in the first DIAGRAM discovery meta-analysis.<sup>18</sup>

### GWAS for continuous glycaemic traits

Initial GWAS interrogated the genetic determinants of T2D as a dichotomous phenotype (disease vs. no disease), rather than examining continuous glycaemic traits. Prior to the development of GWAS, Weedon *et al.* had examined *GCK*, a gene in which rare mutations cause a defect in the rate-limiting enzyme that drives glucose metabolism in the  $\beta$  cell, resulting in MODY2. Using the candidate gene approach, they demonstrated that two common variants, rs1799884 and rs3757840, are associated with fasting glucose surpassing or approaching genome-wide significance ( $1 \times 10^{-9}$  and  $8 \times 10^{-7}$ , respectively).<sup>29</sup>

The association of *GCK* with fasting glucose and the success of categorically driven GWAS primed the field to examine the genetic factors that contribute to the inter-individual variation in glycaemic measures in normoglycaemic subjects. Using a genome-wide approach, *G6PC2* and *MTNR1B* were associated with fasting glucose. It is postulated that *G6PC2* modulates the glycolytic pathway and insulin secretion by dephosphorylating glucose-6-phosphate generated by the  $\beta$ -cell glucose sensor, glucokinase. In a French cohort, carriers of the A allele for rs560887 in the third intron of *G6PC2* had decreased fasting plasma glucose and lower risk of developing mild hyperglycemia in nine-year follow-up.<sup>30</sup> Shortly following the publication of these findings, a study of nondiabetic individuals from Finland and Sardinia reported an association between SNP rs563694 and fasting glucose concentrations. SNP rs563694 is in high LD ( $r^2 = 0.729$ ) with and only 11 kb away from the SNP discovered in the French cohort, rs560887.<sup>31</sup>

In early 2009, three groups concurrently described the influence of *MTNR1B* on quantitative glycaemic traits and further characterized the functional significance of one of these variants, rs10830963.<sup>21–23</sup> *MTNR1B* encodes the melatonin receptor MT2, whose endogenous ligand melatonin is a neurohormone that mediates circadian rhythmicity and appears to influence insulin secretion and glucose levels.<sup>32–34</sup> The largest of the three studies was conducted by the Meta-Analysis of Glucose and Insulin-related traits Consortium

(MAGIC). MAGIC represents an international collaborative effort to combine data from multiple GWAS to identify additional loci that affect glycemic and metabolic traits. This initial study focused on the top association signals from four individual consortia meta-analyses, with each of these consortia providing mutual replication of the top SNPs. This meta-analysis of 10 cohorts, composed of 40,735 individuals, described the association of rs10830963 at *MTNR1B* with fasting glucose ( $P = 3.2 \times 10^{-50}$ ) and T2D ( $P = 3.3 \times 10^{-7}$ ) and confirmed the previously known associations of *GCK* and *G6PC2* with fasting glucose.<sup>21,29–31</sup> In an independent French cohort, a second variant near *MTNR1B* (rs1387153) was associated with T2D, fasting glucose,  $\beta$ -cell function by homeostasis model assessment (HOMA-B<sup>26</sup>), and glycated hemoglobin (HbA<sub>1C</sub>).<sup>22</sup> Lyssenko *et al.* further characterized the physiological ramifications of carrying the risk allele in rs10830963. Carriers of the G allele showed decreased insulin secretion, increasing fasting glucose levels and a higher risk of developing T2D in an average 23.5 year follow-up period, as well as a higher proinsulin-to-insulin ratio.<sup>23</sup>

In follow-up to these initial investigations, a large-scale meta-analysis of all genome-wide data for continuous fasting and 2-h postglucose load traits in nondiabetic participants was performed.<sup>35,36</sup> Twenty-one GWAS of European descent including 46,186 participants analyzed for ~2.5 million genotyped or imputed SNPs coalesced around MAGIC in order to examine fasting glucose, fasting insulin, and fasting indices of  $\beta$ -cell function and insulin resistance (HOMA-B and HOMA-IR respectively).<sup>35</sup> Lead SNPs were replicated among 76,558 additional individuals from 34 additional cohorts (7 of which had undergone GWAS genotyping and contributed *in silico* replication). The joint analysis for fasting glucose uncovered nine new loci including SNPs in or near *ADCY5*, *MADD*, *CRY2*, *ADRA2A*, *FADS1*, *PROX1*, *SLC2A2*, *GLIS3*, and *C2CD4B* and one SNP upstream from *IGF1* in association with fasting insulin and HOMA-IR. The meta-analysis also confirmed prior associations for glycemic traits with SNPs in or near *DGKB*, *TMEM195*, *GCKR*, *G6PC2*, *MTNR1B*, and *GCK*. The 14 fasting glucose-associated loci explained a substantial proportion (10%) of the inherited variation in fasting glucose in the Framingham cohort. In addition, an aggregate genotype score,

constructed by summing the risk variants, showed a difference in fasting glucose of about 7.2 mg/dL between groups with the highest and lowest scores.<sup>35</sup> Contrary to straightforward expectations, only five of the novel fasting glucose-associated SNPs were associated with T2D, which underscores that the mechanism by which glucose is raised, rather than elevated glucose alone, may have contributed to the progression to T2D. The three loci with the largest effect sizes on fasting glucose, *GCPC2*, *MTNR1B*, and *GCK*, were found to have a significant association with HbA<sub>1C</sub> at genome-wide significance thresholds;<sup>37</sup> interestingly, *G6PC2* does not increase risk of T2D. In terms of insulin resistance measures, *IGF1* (rs35767) and *GCKR* (rs780094) were significantly associated with HOMA-IR and fasting insulin.<sup>35</sup>

In order to examine glucose and insulin levels 2 h after an oral glucose tolerance test (OGTT), nine GWAS ( $n = 15,234$ ) were compiled with replication in 17 studies ( $n = 30,620$ ).<sup>36</sup> This study found an association of SNP rs10423928 in *GIPR* (encoding the gastric inhibitory peptide receptor) with glucose levels 2 h after an oral glucose challenge.<sup>36</sup> Gastric inhibitory peptide is an incretin hormone released from the intestine in response to oral glucose, and it stimulates insulin secretion by the pancreatic  $\beta$  cell.<sup>38</sup> GIP response is decreased in T2D.<sup>39</sup> In further analyses supporting the known role of this pathway in glucose regulation, this intronic SNP was associated with a decreased early-phase insulin secretion, decreased insulin response, and lower insulin levels 2 h after an oral, but not intravenous glucose load.

### Genetic variants and physiological parameters

Assessing more detailed physiological parameters influenced by newly discovered genetic variants lends credibility to the original associations, besides offering a better characterization of the mechanism by which these variants might modulate glucose and insulin levels and cause T2D. Consequently, the MAGIC investigators systematically examined the influence of the genetic variants newly associated with fasting glucose, fasting insulin, and/or 2-h glucose with measures that define insulin processing, insulin secretion, and insulin sensitivity.<sup>40</sup> With these measures, loci were categorized into five groups based on their hypothesized mechanism of

action: (1) abnormal insulin processing, (2) higher proinsulin and lower insulin secretion, (3) abnormalities in early insulin secretion, (4) reduced insulin sensitivity, and (5) no obvious effect on insulin processing, secretion, or sensitivity. *MADD* (encoding a death domain-containing adaptor protein that propagates apoptotic signals) was strongly associated with proinsulin levels, but not other parameters. This indicates that the locus is implicated in isolated insulin processing only. *TCF7L2*, *SLC30A8*, *GIPR*, and *C2CD4B* were associated with higher proinsulin and lower insulin secretion. The findings for *TCF7L2* and *SLC30A8* confirmed previous results, which appear biologically plausible given the proposed function of these genes. For instance, genetic variants in *TCF7L2* are thought to adversely affect  $\beta$ -cell responsiveness to incretins and insulin granule exocytosis, which would impair insulin processing and decrease insulin secretion. In turn, variants at *MTNR1B*, *FADS1*, *DGKB*, and *GCK* were only associated with a lower insulinogenic index (which measures the initial phase of glucose-stimulated insulin release, by dividing the change in insulin over the first 30 min of an OGTT by the change in glucose over the same time frame), suggesting that variants in or near these genes impair early insulin secretion. For instance, *FADS1* encodes a key enzyme in metabolism of unsaturated fatty acids, and insulin secretion differs in response to meals with varying fatty acid composition.<sup>41</sup> The lower insulinogenic index caused by genetic variation in *FADS1* indicates that variants in this gene may alter an individual's  $\beta$ -cell response to meals composed of certain fatty acid content. SNPs at *GCKR* and *IGF1* were found to influence insulin sensitivity by multiple methods whose calculation includes measurements beyond fasting glucose and insulin. A functional variant at *GCKR* (encoding the glucokinase regulatory protein) may act by inhibiting glucokinase in the liver, leading to increased hepatic glucose production.<sup>42</sup> Glucose lowering by *IGF1* occurs by the binding of insulin receptors, stimulation of glucose transport into fat and muscle, and thus, lowering glucose and suppression of insulin.<sup>43</sup> Unlike *GCKR*, the influence of the index SNP near *IGF1* on glucose homeostasis is not yet understood. This study confirmed that these variants influence insulin resistance measures that go beyond the fasting state, possibly implicating extra-hepatic tissues in the process.

## What have we learned from GWAS?

In a short span of five years, GWAS investigating the links between genetics and complex traits have transformed our knowledge. As we strive to decipher how to navigate these new roads, it may be instructive to look back at what we have learned to know how to journey forward.

### Box 1. What have we learned from GWAS?

- They illustrate novel pathways.
- The association of a missense mutation rs13266634 in *SLC30A8* (encoding Zn<sup>2+</sup> Transporter, ZnT-8) with type 2 diabetes has highlighted the importance of Zn<sup>2+</sup> transport in the  $\beta$  cell, the variant's influence on insulin packaging and secretion, and this pathway's potential relevance as a drug target.
- New findings point to fundamental biology.
- The intronic SNP rs7903146 in *TCF7L2* is located in an open chromatin site in  $\beta$  cells; its risk T allele is correlated with an increased transcription in human islets and with increased expression in cellular luciferase assays.
- Genetic discoveries support prior epidemiological observations.
- The T allele of rs17550 in *FADS1* is associated with higher fasting glucose, lower HOMA-B, increased LDL and HDL cholesterol, and decreased triglycerides.
- Variants in *HNF1A* are associated with type 2 diabetes, C-reactive protein, coronary artery disease, and LDL cholesterol levels.
- *MTNR1B* and *CRY2* link T2D and circadian rhythms in support of animal and human studies.
- *ADCY5* variants increase fasting and 2-h glucose, as well as risk of type 2 diabetes; they are also associated with lower birth weight.
- Most loci point to the  $\beta$  cell.
- The majority of loci associated with type 2 diabetes point to primary defects in the  $\beta$  cell.
- However, *IRS1* (encoding insulin receptor substrate-1) has been associated with T2D.



- Both *IGF1* and *GCKR* have been associated with measures of insulin resistance (HOMA-IR and fasting insulin) at genome-wide levels of significance.
- Increasing sample size and modifying study design have helped discover variants associated with insulin resistance.
- So far, genetic variants only explain 10% of T2D heritability.
- GWAS have been limited to common variants and primarily populations of European descent.
- In order to find the “missing heritability,” investigators are pursuing fine-mapping around the associated regions, leveraging the 1,000 Genomes project, applying next-generation sequencing, analyzing the Metabochip, using improved informatics for gene  $\times$  gene and gene  $\times$  environment interactions, expanding to nonwhite populations, and incorporating prior biological knowledge to interpret significance of variants.
- Common genetic variants are not yet useful in clinical prediction.
- Although variants may be limited in their ability to predict type 2 diabetes, genetic information may sway a person to change lifestyle behavior that may reduce their risk of developing the disease.
- Genetic prediction may be more useful in younger age groups, before clinical risk factors develop.
- They provide an opportunity for therapeutic intervention and pharmacogenetic clinical trials.
- Pharmacogenetic studies in polygenic diabetes have studied primarily *PPARG*, *KCNJ11*, and *TCF7L2*.
- Two correlated variants at *KCNJ11* and *ABCC8* are associated with sulfonylurea failure and decreased mean fasting glucose on sulfonylurea therapy.
- Carriers of risk variants at *TCF7L2* are more likely to fail sulfonylurea therapy than metformin and more likely to be on insulin therapy rather than diet alone.

### *They illustrate novel pathways*

GWAS have implicated novel pathways in the development of diabetes in humans. One of the earli-

est and most illustrative examples was the discovery that the variant rs13266634, which encodes a R $\rightarrow$ W change at position 325 in the  $\beta$ -cell zinc transporter ZnT-8 (encoded by *SLC30A8*), was strongly associated with T2D (OR 1.26 for the major C allele,  $P = 5.0 \times 10^{-7}$ ) in the first published GWAS for T2D.<sup>13</sup> This association was confirmed by the UK WTCCC ( $P = 10^{-3}$ ),<sup>15</sup> FUSION ( $P = 10^{-5}$ ),<sup>16</sup> and DGI ( $P = 5.3 \times 10^{-8}$ ).<sup>17</sup>

*SLC30A8* is expressed almost exclusively in pancreatic islets with low levels in the cortex and thyroid.<sup>44</sup> Its protein product, ZnT-8, is a zinc transporter localized in secretion vesicle membranes that transports zinc from the cytoplasm into insulin secretory vesicles.<sup>45</sup> Insulin is stored as a hexamer bound to two zinc ions, and ZnT-8 provides zinc to allow for insulin storage and secretion.<sup>46</sup> ZnT-8 thus appears to be a critical component of the final biosynthetic pathway of insulin production and secretion. Interestingly, overexpression of *SLC30A8* in insulinoma cells increases glucose-stimulated insulin secretion.<sup>47</sup> Other *in vitro* studies demonstrate that inflammatory cytokines noted to be elevated in T2D downregulate ZnT-8 in  $\beta$  cells and alter  $\beta$ -cell function.<sup>20,48</sup> The development of a *Slc30a8* null mouse shortly after the GWAS publications further defined how these genetic variants may affect glycemic parameters *in vivo*, although with varying results depending on the mouse model.<sup>49,50</sup> In the mouse models, the size and number of islets as well as insulin sensitivity were comparable between knock-out (KO) mice and wild-type littermates, though the latter showed decreased islet zinc staining. Some of the KO models demonstrated impaired glucose tolerance and fasting glucose depending on age and gender, and one of the mouse models also showed decreased plasma insulin and decreased glucose-stimulated insulin secretion. Thus, though the role of this gene in insulin packaging and storage was known from prior work, GWAS and the experiments spawned by them revealed that coding variation in *SLC30A8* may increase the risk of developing T2D by causing an insulin secretory defect.

### *New findings point to fundamental biology*

While one can easily hypothesize about the biological consequences of missense SNPs, a potential molecular mechanism is not readily apparent for most other associated SNPs, which lie in intronic or

intergenic DNA segments. Therefore, GWAS have reinforced the fundamental quest to characterize the biological relevance of non-coding genomic regions.

One pertinent example concerns the intronic SNP in the gene *TCF7L2*, rs7903146. This SNP has the strongest effect size on T2D demonstrated for a common variant thus far ( $\sim 1.40$  per copy of the risk allele).<sup>51,52</sup> Using a new method called FAIRE-seq (formaldehyde-assisted isolation of regulatory elements coupled with high-throughput sequencing), Gaulton *et al.* identified open chromatin sites in human pancreatic islets.<sup>53</sup> Open chromatin sites are evolutionarily conserved and can be used to tag regions that are bound by regulatory factors.<sup>54</sup> Of 350 SNPs in strong LD with variants associated with T2D or fasting glycemia, 38 SNPs from ten loci (*TCF7L2*, *CDKAL1*, *CDKN2A/CKDN2B*, *IGF2BP2*, *CDC123/CAMK1D*, *THADA*, *FTO*, *SLC30A8*, *HNF1B*, and *G6PC2*) were associated with an open chromatin region in islets. Upon further examination of *TCF7L2*, SNP rs7903146 was localized to an islet-selective open chromatin site; more importantly, the high-risk T allele is associated with a “more open” chromatin state in human islets and greater enhancer activity, as revealed by allele-specific luciferase reporter assays in two  $\beta$ -cell lines.<sup>53</sup> Given that the T allele associates with an increased risk of T2D, the more open chromatin state may imply increased transcription of *TCF7L2*, and thus confirms prior observations that the T allele is correlated with a fivefold increase in *TCF7L2* transcripts in human islets from donors with diabetes when compared to controls.<sup>55</sup> By pairing fundamental biology, such as open chromatin regions indicative of regulatory binding sites, with current genetic association datasets, this study demonstrated a mechanism by which an intronic polymorphism may contribute to the development of a human disease.

### *Genetic discoveries support prior epidemiological observations*

The GWAS method has not only led us to discover new or confirm known biology, but also underscored prior epidemiological observations by providing potential genetic links between lipid dysregulation and glycemia (*FADS1*, *GCKR*, *HNF1A*), circadian rhythmicity and metabolic derangements (*MTNR1B*, *CRY2*), and low birth weight with subsequent T2D risk (*ADCY5*). A potential connection

between genetic determinants of T2D that simultaneously lower cancer risk (*HNF1B*, *JAZF1*) is discussed at length elsewhere.<sup>56</sup>

Fatty acid desaturases (FADS) convert polyunsaturated fatty acids into cell signaling metabolites that in turn can affect circulating lipid levels.<sup>57</sup> Different GWAS have associated *FADS1* with fasting glucose and HOMA-B (rs174550,  $P = 2 \times 10^{-15}$ , and  $5 \times 10^{-13}$ , respectively), HDL cholesterol and triglyceride levels (rs174547,  $P = 2 \times 10^{-12}$ , and  $2 \times 10^{-14}$ , respectively), and LDL cholesterol levels (rs174546,  $P = 1 \times 10^{-7}$ ).<sup>35,57,58</sup> All of these SNPs are in complete LD ( $r^2 = 1$ ) and therefore serve as perfect proxies for each other. In a large meta-analysis of glycemic traits, the T allele at rs174550 was associated with higher fasting glucose, lower HOMA-B, increased HDL and LDL cholesterol, and decreased triglyceride levels.<sup>35</sup> Classically, T2D and impaired fasting glucose have been clustered with dyslipidemia in the metabolic syndrome; the discovery of a genetic variant that is associated with both measures provides support for this clinical correlation. More recently, *FADS1* was found to be associated with a statin-induced response in triglyceride and HDL cholesterol levels ( $P = 2.6 \times 10^{-6}$  and  $6.8 \times 10^{-6}$ , respectively).<sup>59</sup> Given that unsaturated fatty acids influence insulin secretion, this genetic variant can tightly link hyperglycemia and dyslipidemia.<sup>41</sup> Although the same SNP appears to influence various metabolic parameters, each of these measures could actually be influenced by different SNPs in high LD with the index SNP; elucidation of molecular mechanism requires functional experiments.

*HNF1A* (encoding hepatocyte nuclear factor-1 $\alpha$ ) illustrates a locus that contributes to the inheritance of both a rare Mendelian disorder and common polygenic disease, while also having pleiotropic effects. Classically, MODY3 is caused by a mutation in the *HNF1A* gene, which impairs the dimerization of this transcription factor and thus promotes metabolic dysregulation resulting in diabetes mellitus.<sup>60</sup> Like *FADS1*, SNPs in the *HNF1A* region have been associated with multiple related traits including T2D ( $P = 2 \times 10^{-8}$ ), C-reactive protein ( $P = 2 \times 10^{-8}$  to  $1 \times 10^{-30}$ ), coronary artery disease ( $P = 5 \times 10^{-7}$ ), and LDL cholesterol ( $P = 2 \times 10^{-8}$ ).<sup>57,61–65</sup> In terms of lipid metabolism and inflammation, *HNF1A* regulates numerous genes involving lipoprotein metabolism

and inflammatory markers in the liver.<sup>66</sup> These genetic variants create another molecular link between related metabolic derangements.

*MTNR1B* is an equally compelling story that helps provide a genetic link between circadian rhythmicity and glucose metabolism. As discussed above, *MTNR1B* variants have been associated with glycemic quantitative traits including T2D, fasting glucose, HOMA-B, and HbA<sub>1C</sub>.<sup>21–23</sup> A second circadian gene, *CRY2*, has also been implicated in regulating fasting glucose in nondiabetic individuals.<sup>35</sup> Epidemiological and physiological studies in humans, as well as experimental work in animals, have established a clear relationship between alterations of the circadian system and metabolic derangements. Homozygous mutant *Clock* mice, which are deficient in a key circadian rhythm transcription factor, were noted to have a variation in nighttime activity, in addition to metabolic disturbances including hyperphagia, obesity, hyperlipidemia, hepatic steatosis, hyperglycemia, and hypoinsulinemia.<sup>67</sup> Furthermore, wild-type islets appear to possess a self-sustained oscillation of a circadian gene (*Per2*) and circadian transcription factors (*CLOCK* and *BMAL1*) that are not seen in the islets of *Clock* null mice. *In vivo* studies in *Clock* mutant mice and *in vitro* studies in islets from whole-body *Clock* and *Bmal1* mutant mice, as well as islets from pancreas-specific *Bmal1* mutant mice, demonstrated impairment in insulin secretion.<sup>68</sup> In order to explain the higher glucose intolerance noted in the morning versus night, Boden *et al.* studied normal human volunteers undergoing a hyperglycemic clamp titrated to various levels of glucose control for 24 h. By measuring insulin secretion, the investigators demonstrated that, particularly at higher glucose levels, insulin is secreted in a circadian rhythm.<sup>32</sup> In addition, people with T2D demonstrate a decreased diurnal serum melatonin level.<sup>34</sup> The link between genetic variants in *MTNR1B* and *CRY2* with T2D and glycemic measures lends further credence to the relationships observed in these animal and human studies.

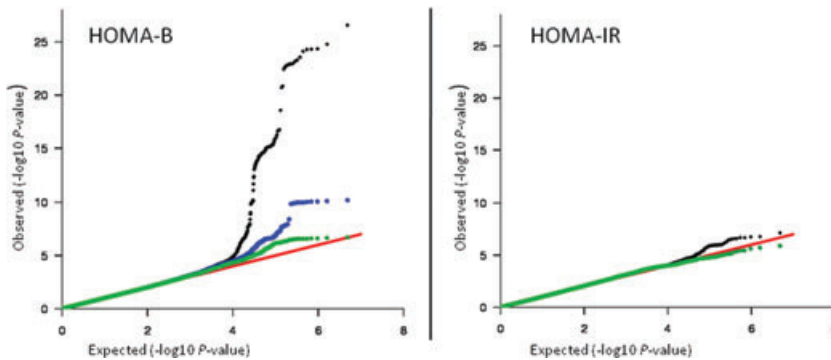
Finally, genetic variants in *ADCY5* (encoding adenylate cyclase type 5) have been robustly associated with fasting glucose, 2-h glucose post-OGTT, and T2D.<sup>35,36</sup> In recent work confirmed by an independent group, the same locus has been associated with low birth weight.<sup>69,70</sup> This finding provides another molecular connection between the observed

predictive power of reduced size at birth with regard to future metabolic derangements, including risk of T2D.<sup>71</sup>

### *Most loci point to the $\beta$ -cell*

It is now clear that the risk of developing T2D is a combination of genetic risk for  $\beta$ -cell dysfunction superimposed on genetic and environmental factors (e.g. obesity, Western diet, sedentary lifestyle) that promote insulin resistance. Recent genetic discoveries have identified numerous variants that appear to influence insulin secretion rather than insulin resistance. For example, risk variants of *CDKAL1* were associated with insulin secretion defects and impaired insulin response to both oral and IV glucose tolerance tests (IVGTT).<sup>14,72,73</sup> Variants in *SLC30A8* were associated with insulin secretion defects following OGTT.<sup>14,74</sup> *HHEX* variants linked with impaired  $\beta$ -cell function in response to OGTT and hyperinsulinemic-euglycemic clamp.<sup>72,74,75</sup> Finally, *CDKN2A/B* variants have been associated with impaired glucose-induced insulin release in healthy subjects.<sup>75</sup> Of the ten novel variants discovered in association with fasting glycemic traits in a large meta-analysis, nine were associated with fasting glucose/HOMA-B and only one was associated with fasting insulin/HOMA-IR, both measures of insulin resistance.<sup>35</sup>

There are several potential reasons for the dearth of common genetic variants related to insulin resistance found by the GWAS approach. These include study design, different heritability estimates for each of the traits, and the allelic frequency spectrum of these variants. One of the five initial high-density GWAS selected diabetic cases with BMI < 30 kg/m<sup>2</sup>, thus excluding obese individuals,<sup>13</sup> while another matched cases and controls for BMI.<sup>17</sup> These deliberate study designs were chosen to maximize the likelihood of detecting variants that increase T2D risk directly, rather than through the mediation of adiposity. Because loci that achieve genome-wide significance result from the meta-analyses of various scans, some loci that truly cause T2D via insulin resistance may show a trend toward association that fails to reach a threshold that marks them for follow-up. In this way, the studies may not identify loci associated with insulin resistance related to adiposity, but rather, detect loci related only to  $\beta$ -cell function. This issue has become less of a concern as more recent meta-analyses incorporate



**Figure 3.** Quantile–quantile plots for HOMA-B and HOMA-IR in a large meta-analysis. The quantile–quantile (QQ) plot illustrates the observed  $P$  values in the distribution versus those that were expected under the null hypothesis of no association. The null distribution is depicted by the red diagonal line. The entire distribution of  $P$  values is in black, the exclusion of the ten newly discovered loci (*DGKB-TMEM195*, *ADCY5*, *MADD*, *ADRA2A*, *FADS1*, *CRY2*, *SLC2A2*, *GLIS3*, *PROX1*, and *C2CD4B*) in green, and the exclusion of the four genome-wide significant fasting glucose-associated loci reported previously (*GCK*, *GCKR*, *G6PC2*, and *MTNR1B*) in blue. The HOMA-B  $P$  values have a larger deviation from expected compared to HOMA-IR, and, therefore, genetic associations with HOMA-B are more likely to be detected than with HOMA-IR. Figure adapted from Ref. 35.

larger numbers of discovery cohorts that are not ascertained by BMI.

Investigators may bypass this limitation by performing GWAS for insulin resistance as a quantitative trait, rather than basing the discovery of insulin resistance genes in scans for T2D as a categorical trait. One measure that is readily available in large population cohorts is HOMA-IR. However, a meta-analysis of GWAS cohorts in which this phenotype was available identified many more loci influencing  $\beta$ -cell function than those regulating HOMA-IR, suggesting that other reasons exist for this observed discrepancy.

The degree to which each of the two traits is heritable may also contribute to the difficulties in identifying insulin resistance genes. The insulinogenic index is, on average, 10% more heritable than HOMA-IR.<sup>76</sup> While HOMA-IR still displays substantial heritability (0.44 among the Framingham Offspring population<sup>77</sup>), there may be better insulin resistant phenotypes to examine in order to identify genetic associations. For example, in the Insulin Resistance and Atherosclerosis Study Family Study, the insulin sensitivity index (determined from IVGTT) was twice as heritable as HOMA-IR.<sup>78</sup> This suggests that the insulin sensitivity index may be a more robust trait to examine when assessing the genetic contribution of insulin resistance; this must be balanced with the difficulty in performing the necessary physiology experiments in a sufficiently high number of individuals.

The distribution of  $P$  values in the similarly powered GWAS for HOMA-B and HOMA-IR (both derived from fasting glucose and insulin) suggests a different genetic architecture for the two traits.<sup>35</sup> In those meta-analyses, there were many more  $P$  values that deviated from the null expectation at the top of the HOMA-B distribution than there were for HOMA-IR (Fig. 3). For instance, insulin resistance variants may be fewer in number, rarer in frequency, or have a lower effect size. In addition, there may be a stronger impact of environmental interactions (e.g. diet, physical activity) with the genetic background. Exploring the lower range of the allele frequency spectrum (which requires denser genotyping arrays), integrating environmental factors in the analyses, and increasing the sample size could potentially improve the power to detect genetic variants related to insulin resistance.

Nevertheless, several convincing associations with surrogate measures of insulin resistance (HOMA-IR and fasting insulin) have been identified, in part due to larger sample sizes and a novel study design. *IRS1* is a highly attractive candidate gene for the development of insulin resistance. Rung *et al.* demonstrated that SNP rs2943641, 502 kb upstream of *IRS1*, is associated with T2D, fasting insulin, and HOMA-IR. In further functional analyses, the risk allele was associated with reduced levels of *IRS1* protein expression and decreased downstream effects noted by phosphatidylinositol-3-OH kinase (PI(3)K) activity (which is activated by

IRS1 binding to IGF-1 receptors) in human skeletal muscle biopsies. Such functional experiments were critical in establishing that this variant, located at a non-trivial distance from *IRS1*, likely acts through *IRS1* itself. These findings confirm what is already known biologically about the action of IRS1 in mice. Null *Irs1* mice lack PI(3)K activity and demonstrate impaired glucose tolerance and insulin resistance.<sup>79</sup>

The most recent meta-analysis for fasting glycemic traits uncovered the variant rs35767 1.2 kb upstream of *IGF1* as associated with fasting insulin and HOMA-IR.<sup>35</sup> *IGF1* is another excellent biological candidate for insulin resistance: it encodes the insulin-like growth factor-1, which has been shown to bind insulin receptors and enhance glucose transport in adipose and muscle tissue, while inhibiting hepatic glucose production. Both a null *IGF1* mouse and a human with *IGF1* gene deletion have resulted in insulin resistance that improves with IGF1 therapy.<sup>43</sup> Therefore, it seems intuitive that genetic variation in *IGF1* may perturb the expression or function of IGF1, and thereby increase insulin resistance.

In the same large meta-analysis, *GCKR* was associated with insulin resistance at genome-wide significance levels. Originally, the T allele of rs780094, an intronic SNP in the *GCKR* gene, was found to be associated with higher triglyceride levels ( $P = 3.7 \times 10^{-8}$ ), and carriers of the T allele were noted to have a trend toward lower glucose, decreased insulin resistance (HOMA-IR), and decreased T2D risk.<sup>15</sup> Two follow-up studies confirmed these results.<sup>80,81</sup> The directions of these trends (higher triglyceride levels but lower fasting glucose and lower insulin resistance) may be counterintuitive with what we have seen in human epidemiologic studies. This may be explained by the inherent biological mechanism of *GCKR*, where the functional variant may downregulate gluconeogenesis, increase VLDL-triglyceride synthesis, and also upregulate glucose use.<sup>42,81</sup>

It is clear that as far as the genetic architecture of insulin resistance, we have only visualized the tip of the iceberg. A potentially large proportion of the genetic foundation of T2D remains hidden owing to our difficulties in discovering insulin resistance-related genes. With well-powered samples, more refined phenotyping, assessment of gene-environment interactions, a wider range of BMI,

and novel study designs, this genetic information may soon rise to the surface.

### *Genetic variants identified only explain ~10% of T2D heritability*

T2D GWAS have been successful in identifying specific loci that contribute to the causation of the complex disease, but only roughly 10% of the heritability can be accounted for by these variants, suggesting that much remains to be discovered. In the search for the “missing heritability,” firstly the accuracy of original heritability estimates needs to be considered. The familial component caused by a shared early uterine and postnatal environment and latent epigenetic changes (inherited changes in gene expression that are not caused by genetic sequence) may produce a contribution to heritability that is not removed sufficiently by these estimates.<sup>82</sup> In addition, GWAS use commercially available genotyping platforms whose early-generation arrays fail to adequately capture nearly 20% of common SNPs, structural variants (such as copy number variants), and variants unique to non-European populations.<sup>83</sup> Improving upon these limitations of first-generation genotyping platforms will undoubtedly uncover other loci contributing to the disease. In addition, since a common SNP represents a large segment of genetic material, closer scrutiny of these areas by fine mapping for rare variants is required to find the potential “causal” gene that could have a larger effect than that conferred by the index variant. Through purifying natural selection, high-penetrance deleterious variants should be selectively removed from the population, so we would not expect to find common variants contributing large effect sizes to common disease, but may find large effect sizes in less common variants. Whole-genome sequencing with less costly high-throughput methods have been developed to discover rarer variants in the human genome. The 1,000 Genomes Project ([www.1000genomes.org](http://www.1000genomes.org)) is an international initiative designed to catalog all variants with minor allele frequencies greater than 1% of at least 1,000 genomes, thus pushing the envelope of captured shared variation. Initial pilot analyses have successfully identified more than 9 million new polymorphisms, many insertions/deletions, and some large structural variants.<sup>84</sup>

The Metachip is another effort to develop a single custom-made chip encompassing 200,000 SNPs

culled from deeper layers of the *P*-value distribution in GWAS for cardiovascular disease, obesity, T2D, and related traits. As the Metabochip is deployed across thousands of patients, new and rare variants whose *P* values are near-the-top but below threshold will be examined in association with these phenotypes. Researchers are also incorporating novel bioinformatics to examine nonadditive gene  $\times$  gene interactions (commonly referred as epistasis). For example, two variants may individually affect genetic risk of T2D only mildly, but together they could increase the risk significantly. In addition, examining gene  $\times$  environment interactions might lead to insights into biological pathways and guide researchers to novel genes that act synergistically with environmental factors, such as physical activity or dietary composition.<sup>85</sup> If successful, this strategy may explain why some genetic associations do not replicate across all populations, hinting to different environmental exposures. Finally, incorporating the prior probabilities afforded by the mechanistic knowledge garnered through functional experiments in cellular or animal models, investigators may rescue variants that did not make the list of associated loci on purely statistical grounds.

#### *Common genetic variants are not yet useful for clinical prediction*

In practical terms, genetic information has been expected to enable clinicians to predict an individual's risk of developing disease. However, thus far the clinical usefulness of genetic information has been limited. Meigs *et al.* examined 18 loci associated with T2D in the Framingham Offspring Study in a genotype score in association with T2D incidence, and generated C statistics to determine the individual's future risk of diabetes.<sup>86</sup> While the genotype score predicted diabetes better than gender alone, when adjusted for other easily obtainable clinical parameters (age, sex, family history, BMI, fasting glucose level, systolic blood pressure, HDL cholesterol, and triglyceride levels), predictive power did not improve significantly. A similar study in a larger Swedish cohort examined 16 SNPs associated with T2D and generated largely comparable results.<sup>87</sup> In subset analyses, both studies highlighted that younger patients may benefit more from genetic testing before they manifest clinical characteristics of the disease. An updated age-stratified analysis re-

examined the genotype score including 40 SNPs associated with T2D in Framingham. The researchers noted that the new genotype score improved the ability to predict the onset of T2D in participants younger than 50 years. Diabetes risk prediction using the genetic variants still needs to be confirmed by other studies, but appears potentially promising in a younger population when medical and lifestyle intervention may be the most feasible and beneficial.<sup>88</sup>

An alternative clinical use of genetic information is determining whether genetic information may affect a person's behavior and thus modify the risk of developing diabetes. The Chinese Da Qing Diabetes Prevention Study, the Finnish Diabetes Prevention Study, the U.S. Diabetes Prevention Program (DPP), and the Indian Diabetes Prevention Program showed that behavioral modification can reduce one's risk of progressing to diabetes.<sup>89–92</sup> For instance, the DPP demonstrated that lifestyle modification led to a 58% reduced incidence of the development of diabetes compared to placebo.<sup>92</sup> In further genetic investigation in the DPP, lifestyle modification attenuated the risk of diabetes conferred by the risk variant at *TCF7L2*: while the placebo participants who carried the homozygous risk genotype at rs7903146 (TT) had 80% higher risk of diabetes development, those randomized to the lifestyle intervention group had no increased risk despite carrying the risk genotype, indicating that lifestyle intervention may trump the genetic risk conferred by this variant.<sup>93</sup> This argues that if genetic information could motivate a high-risk individual to change his/her behavior, it may help prevent the eventual onset of disease. However, whether reassuring genetic information may be counterproductive in demotivating those at lower genetic risk from engaging in healthy behaviors should be empirically tested.

The concept of disease-risk information changing one's behavior was examined among 150 nondiabetic primary care patients.<sup>94</sup> Those with an accurate perception of their higher risk of diabetes (as documented by the Framingham Heart Study diabetes risk score) did not intend to modify their lifestyle any more than patients with low perceived risk. Whether genetic information (rather than perceived or measured risk factors) will have greater influence is being tested. Grant *et al.* examined

how genetic information might influence an individual's motivation for lifestyle change.<sup>95</sup> After questioning patients regarding how they would respond to information containing their genetic risk of T2D development, 71% of those surveyed indicated that such testing would motivate them to adopt lifestyle change. On the contrary, only 1.3% of the patients reported they would be less motivated by genetic information indicating they were at low risk. Therefore, despite the limited ability of genetic testing to predict T2D when adjusted for common risk factors, genetic information may be more powerful in persuading a patient to change his or her behavior. Further studies, including more surveyed individuals and prospective studies, are needed to determine if the person's behavior would in fact change.

#### *Genetic findings provide an opportunity for therapeutic intervention and pharmacogenetic clinical trials*

The underlying biological mechanisms that GWAS have unveiled could be targets for new pharmacotherapy. For example, prior biological knowledge regarding *SLC30A8* and new genetic evidence have drawn our attention to a possible therapeutic target. A pharmacological agent that enhances the function of this transporter could hypothetically increase insulin secretion and improve  $\beta$ -cell function. Given the novelty of the GWAS findings, there has yet to be a therapeutic that has specifically evolved from these discoveries, but it seems that there is a wide breadth of possibilities on the horizon.

Studies to examine the pharmacogenetics of T2D have been encouraged by findings of variable pharmacotherapy response among individuals with monogenic forms of diabetes. One of the most impressive stories lies in the discovery of activating mutations in *KCNJ11* (encoding the Kir6.2 subunit of the ATP-sensitive potassium channel), which causes the channels to remain open in the presence of glucose, thereby reducing insulin secretion and giving rise to permanent neonatal diabetes (PNDM).<sup>96</sup> Infants carrying these mutations present with marked hyperglycemia or diabetic ketoacidosis at less than six months of age. Three individuals harboring mutations in *KCNJ11* were successfully transitioned from insulin to a sulfonylurea with an improvement in glycemic control compared to insulin ther-

apy.<sup>97</sup> These findings were later confirmed by larger study populations of patients with PNDM.<sup>96,98</sup> Similarly, the enhanced efficacy of sulfonylureas in patients with *MODY3* was described with a fivefold greater response to sulfonylurea versus metformin therapy.<sup>99</sup>

Preliminary pharmacogenetic studies in polygenic diabetes have been initiated. These studies have primarily involved studying *PPARG*, *KCNJ11*, and *TCF7L2*, whose variants have been associated with T2D at genome-wide levels of significance (as previously mentioned, the first two encode known targets of antidiabetic medications).

Given its therapeutic relevance in PNDM and *MODY3*, the influence of *KCNJ11* variants has been examined in response to sulfonylurea treatment in T2D. Sesti *et al.* conducted a prospective trial of 524 European T2D patients who were treated with a progressive antidiabetic medication regimen, which escalated with each medication failure. They started with glibenclamide, followed by metformin and then insulin. Carriers of the T2D risk K allele had a relative risk of 1.45 compared to E23E homozygotes for sulfonylurea failure.<sup>100</sup> Feng *et al.* found an association of rs757110 (a SNP in the adjacent gene *ABCC8*, which encodes the sulfonylurea receptor SUR1) with response to sulfonylurea treatment in Chinese patients with T2D, measured by a decrease in mean fasting glucose.<sup>101</sup> The missense mutation at rs757110, S1369A, is in high LD ( $r^2 = 0.87$ ) with the E23K polymorphism, such that they are statistically indistinguishable. In contrast to Sesti *et al.*, in the Chinese study, carriers of the risk genotype showed greater responsiveness to gliclazide, which is consistent with the PNDM findings; whether such responsiveness also heralds a greater risk of long-term sulfonylurea failure must be tested in longer pharmacogenetic trials.

*PPARG* is an attractive genetic target because it is the known drug target of thiazolidinediones such as troglitazone, rosiglitazone, and pioglitazone.<sup>102</sup> Initial studies reporting that genotype at the index variant in *PPARG* (rs1801282, P12A) or at other nearby variants predict a better response to troglitazone therapy<sup>103,104</sup> have not been substantiated.<sup>105–107</sup>

Finally, *TCF7L2* has also been under investigation with regards to response to antidiabetic medication, despite its mechanism not being as well understood as that of *KCNJ11/ABCC8* or *PPARG*. The risk



variant in this gene confers a defect in insulin secretion and possibly affects GLP-1 metabolism.<sup>108</sup> The Go-DARTS study group has examined the variation in medication response based on *TCF7L2* genotype in two settings. First, they noticed an overrepresentation of the T risk allele in people with T2D treated with insulin versus diet therapy alone. In a second study, they demonstrated carriers of the risk *TCF7L2* variants were more likely to fail sulfonylurea therapy than metformin.<sup>109,110</sup>

### Which road do we take from here?

GWAS have led the way toward an exciting journey into understanding the genetic architecture of T2D and related traits. Critical building blocks have been established. With the development of HapMap and 1,000 Genomes, we have the ability to use imputation to combine different genotyping platforms and continue international collaborations involving multiple cohorts. Advancing technology has enabled large-scale genotyping with high quality data. Innovative statistical analyses, often publically available, have allowed the field to analyze this growing wealth of data. Strategies to identify causal variants include deploying next-generation sequencing techniques, examining the less common allelic spectrum, exploiting genetic pleiotropy, fine-mapping loci with biological relevance, and applying biological insight to the discovered associations. Ultimately, investigators strive to translate this genetic knowledge to the clinician's bedside. Perhaps most importantly, the collaborative spirit that has permeated the field may be the most significant achievement catalyzing our enhanced understanding of the genetics underlying this complex disease.

#### Box 2. Which road do we take from here?

- Using next-generation sequencing techniques to discover less common and rare variation.
- Fine-mapping and applying biological insight to the discovered associations to determine the causal variant.
- Exploiting genetic pleiotropy to understand how these genetic variants link common diseases.
- Continuing to develop informatics methods that examine epistasis and gene × environment interaction.

- Examining populations of non-European descent.
- Translating genetic knowledge to the clinician's bedside with targeted treatment algorithms and risk assessment tools that may influence patient's behavior based on risk.
- Leveraging the welcome collaborative spirit that has permeated the field to catalyze large international studies to enhance our understanding of the genetics of T2D.

### Acknowledgments

L.K.B. is supported by NIH training grant 5 T32 DK007028-35. J.C.F. is supported by the Massachusetts General Hospital and a Clinical Scientist Development Award by the Doris Duke Charitable Foundation. We thank Erik Billings for help in adapting Figure 1. We thank the members of the Florez research group and the larger diabetes genetics community for fruitful and engaging discussions.

### Conflicts of interest

The authors declare no conflicts of interest.

### References

1. Vaxillaire, M. & P. Froguel. 2006. Genetic basis of maturity-onset diabetes of the young. *Endocrinol. Metab. Clin. North Am.* **35**: 371–384.
2. Reynisdottir, I. *et al.* 2003. Localization of a susceptibility gene for type 2 diabetes to chromosome 5q34-q35.2. *Am. J. Hum. Genet.* **73**: 323–335.
3. Grant, S.F. *et al.* 2006. Variant of transcription factor 7-like 2 (*TCF7L2*) gene confers risk of type 2 diabetes. *Nat. Genet.* **38**: 320–323.
4. Altshuler, D. *et al.* 2000. The common PPARgamma Prol2Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat. Genet.* **26**: 76–80.
5. Gloyn, A.L. *et al.* 2003. 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.
6. Florez, J.C. *et al.* 2004. Haplotype structure and genotype-phenotype correlations of the sulfonylurea receptor and the islet ATP-sensitive potassium channel gene region. *Diabetes* **53**: 1360–1368.
7. Lander, E.S. *et al.* 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.
8. Venter, J.C. *et al.* 2001. The sequence of the human genome. *Science* **291**: 1304–1351.
9. The International HapMap Consortium. 2003. The International HapMap Project. *Nature* **426**: 789–796.



10. Frazer, K.A. *et al.* 2007. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**: 851–861.
11. Sachidanandam, R. *et al.* 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**: 928–933.
12. The International HapMap Consortium. 2005. A haplotype map of the human genome. *Nature* **437**: 1299–1320.
13. Sladek, R. *et al.* 2007. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* **445**: 881–885.
14. Steinthorsdottir, V. *et al.* 2007. A variant in *CDKAL1* influences insulin response and risk of type 2 diabetes. *Nat. Genet.* **39**: 770–775.
15. Zeggini, E. *et al.* 2007. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* **316**: 1336–1341.
16. Scott, L.J. *et al.* 2007. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316**: 1341–1345.
17. Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research. 2007. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316**: 1331–1336.
18. Zeggini, E. *et al.* 2008. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat. Genet.* **40**: 638–645.
19. Pe'er, I., R. Yelensky, D. Altshuler & M.J. Daly. 2008. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet. Epidemiol.* **32**: 381–385.
20. Egeffjord, L. *et al.* 2009. Zinc transporter gene expression is regulated by pro-inflammatory cytokines: a potential role for zinc transporters in beta-cell apoptosis? *BMC Endocr. Disord.* **9**: 7–17.
21. Prokopenko, I. *et al.* 2009. Variants in *MTNR1B* influence fasting glucose levels. *Nat. Genet.* **41**: 77–81.
22. Bouatia-Naji, N. *et al.* 2009. A variant near *MTNR1B* is associated with increased fasting plasma glucose levels and type 2 diabetes risk. *Nat. Genet.* **41**: 89–94.
23. Lyssenko, V. *et al.* 2009. Common variant in *MTNR1B* associated with increased risk of type 2 diabetes and impaired early insulin secretion. *Nat. Genet.* **41**: 82–88.
24. Yasuda, K. *et al.* 2008. Variants in *KCNQ1* are associated with susceptibility to type 2 diabetes mellitus. *Nat. Genet.* **40**: 1092–1097.
25. Unoki, H. *et al.* 2008. SNPs in *KCNQ1* are associated with susceptibility to type 2 diabetes in East Asian and European populations. *Nat. Genet.* **40**: 1098–1102.
26. Matthews, D.R. *et al.* 1985. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* **28**: 412–419.
27. Rung, J. *et al.* 2009. Genetic variant near *IRS1* is associated with type 2 diabetes, insulin resistance and hyperinsulinemia. *Nat. Genet.* **41**: 1110–1115.
28. Kassem, S.A. *et al.* 2001. p57(KIP2) expression in normal islet cells and in hyperinsulinism of infancy. *Diabetes* **50**: 2763–2769.
29. Weedon, M.N. *et al.* 2006. A common haplotype of the glucokinase gene alters fasting glucose and birth weight: association in six studies and population-genetics analyses. *Am. J. Hum. Genet.* **79**: 991–1001.
30. Bouatia-Naji, N. *et al.* 2008. A polymorphism within the *G6PC2* gene is associated with fasting plasma glucose levels. *Science* **320**: 1085–1088.
31. Chen, W.M. *et al.* 2008. Variations in the *G6PC2/ABCB11* genomic region are associated with fasting glucose levels. *J. Clin. Invest.* **118**: 2620–2628.
32. Boden, G., J. Ruiz, J.L. Urbain & X. Chen. 1996. Evidence for a circadian rhythm of insulin secretion. *Am. J. Physiol.* **271**: E246–252.
33. Van Cauter, E. 1998. Putative roles of melatonin in glucose regulation. *Therapie* **53**: 467–472.
34. Peschke, E. *et al.* 2006. Diabetic Goto Kakizaki rats as well as type 2 diabetic patients show a decreased diurnal serum melatonin level and an increased pancreatic melatonin-receptor status. *J. Pineal. Res.* **40**: 135–143.
35. Dupuis, J. *et al.* 2010. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat. Genet.* **42**: 105–116.
36. Saxena, R. *et al.* 2010. Genetic variation in *GIPR* influences the glucose and insulin responses to an oral glucose challenge. *Nat. Genet.* **42**: 142–148.
37. Soranzo, N. *et al.* 2010. Common variants at ten genomic loci influence hemoglobin A1C levels via glycemic and non-glycemic pathways. *Diabetes*. 2010 September 21 [Epub ahead of print].
38. Gautier, J.F., S.P. Choukem & J. Girard. 2008. Physiology of incretins (GIP and GLP-1) and abnormalities in type 2 diabetes. *Diabetes Metab.* **34**(Suppl 2): S65–S72.
39. Nauck, M.A. *et al.* 1986. Incretin effects of increasing glucose loads in man calculated from venous insulin and C-peptide responses. *J. Clin. Endocrinol. Metab.* **63**: 492–498.
40. Ingelsson, E. *et al.* 2010. Detailed physiologic characterization reveals diverse mechanisms for novel genetic Loci regulating glucose and insulin metabolism in humans. *Diabetes* **59**: 1266–1275.
41. Xiao, C., A. Giacca, A. Carpentier & G.F. Lewis. 2006. Differential effects of monounsaturated, polyunsaturated and saturated fat ingestion on glucose-stimulated insulin secretion, sensitivity and clearance in overweight and obese, non-diabetic humans. *Diabetologia* **49**: 1371–1379.
42. Beer, N.L. *et al.* 2009. The P446L variant in *GCKR* associated with fasting plasma glucose and triglyceride levels exerts its effect through increased glucokinase activity in liver. *Hum. Mol. Genet.* **18**: 4081–4088.
43. Clemmons, D.R. 2004. Role of insulin-like growth factor in maintaining normal glucose homeostasis. *Horm. Res.* **62**(Suppl 1): 77–82.
44. Murgia, C. *et al.* 2009. Diabetes-linked zinc transporter ZnT8 is a homodimeric protein expressed by distinct rodent endocrine cell types in the pancreas and other glands. *Nutr. Metab. Cardiovasc. Dis.* **19**: 431–439.
45. Chimienti, F., S. Devergnas, A. Favier & M. Seve. 2004. Identification and cloning of a beta-cell-specific zinc transporter, ZnT-8, localized into insulin secretory granules. *Diabetes* **53**: 2330–2337.

46. Chimienti, F., A. Favier & M. Seve. 2005. ZnT-8, a pancreatic beta-cell-specific zinc transporter. *Biometals* **18**: 313–317.
47. Chimienti, F. *et al.* 2006. In vivo expression and functional characterization of the zinc transporter ZnT8 in glucose-induced insulin secretion. *J. Cell. Sci.* **119**: 4199–4206.
48. El Muayed, M. *et al.* 2010. Acute cytokine-mediated down-regulation of the zinc transporter ZnT8 alters pancreatic beta-cell function. *J. Endocrinol.* **206**: 159–169.
49. Pound, L.D. *et al.* 2009. Deletion of the mouse *Slc30a8* gene encoding zinc transporter-8 results in impaired insulin secretion. *Biochem. J.* **421**: 371–376.
50. Nicolson, T.J. *et al.* 2009. Insulin storage and glucose homeostasis in mice null for the granule zinc transporter ZnT8 and studies of the type 2 diabetes-associated variants. *Diabetes* **58**: 2070–2083.
51. Cauchi, S. *et al.* 2007. *TCF7L2* is reproducibly associated with type 2 diabetes in various ethnic groups: a global meta-analysis. *J. Mol. Med.* **85**: 777–782.
52. Florez, J.C. 2007. The new type 2 diabetes gene *TCF7L2*. *Curr. Opin. Clin. Nutr. Metab. Care.* **10**: 391–396.
53. Gaulton, K.J. *et al.* 2010. A map of open chromatin in human pancreatic islets. *Nat. Genet.* **42**: 255–259.
54. Wallrath, L.L., Q. Lu, H. Granok & S.C. Elgin. 1994. Architectural variations of inducible eukaryotic promoters: preset and remodeling chromatin structures. *Bioessays* **16**: 165–170.
55. Weedon, M.N. *et al.* 2007. A common variant of *HMG2* is associated with adult and childhood height in the general population. *Nat. Genet.* **39**: 1245–1250.
56. Frayling, T.M., H. Colhoun & J.C. Florez. 2008. A genetic link between type 2 diabetes and prostate cancer. *Diabetologia* **51**: 1757–1760.
57. Kathiresan, S. *et al.* 2009. Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat. Genet.* **41**: 56–65.
58. Sabatti, C. *et al.* 2009. Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat. Genet.* **41**: 35–46.
59. Barber, M.J. *et al.* 2010. Genome-wide association of lipid-lowering response to statins in combined study populations. *PLoS One*. **5**: e9763.
60. Hua, Q.X. *et al.* 2000. Diabetes-associated mutations in a beta-cell transcription factor destabilize an antiparallel "mini-zipper" in a dimerization interface. *Proc. Natl. Acad. Sci. USA.* **97**: 1999–2004.
61. Erdmann, J. *et al.* 2009. New susceptibility locus for coronary artery disease on chromosome 3q22.3. *Nat. Genet.* **41**: 280–282.
62. Elliott, P. *et al.* 2009. Genetic Loci associated with C-reactive protein levels and risk of coronary heart disease. *JAMA.* **302**: 37–48.
63. Reiner, A.P. *et al.* 2008. Polymorphisms of the *HNFI1A* gene encoding hepatocyte nuclear factor-1 alpha are associated with C-reactive protein. *Am. J. Hum. Genet.* **82**: 1193–1201.
64. Ridker, P.M. *et al.* 2008. Loci related to metabolic-syndrome pathways including *LEPR*, *HNFI1A*, *IL6R*, and *GCKR* associate with plasma C-reactive protein: the Women's Genome Health Study. *Am. J. Hum. Genet.* **82**: 1185–1192.
65. Voight, B.F. *et al.* 2010. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet.* **42**: 579–589.
66. Odom, D.T. *et al.* 2004. Control of pancreas and liver gene expression by HNF transcription factors. *Science* **303**: 1378–1381.
67. Turek, F.W. *et al.* 2005. Obesity and metabolic syndrome in circadian Clock mutant mice. *Science* **308**: 1043–1045.
68. Marcheva, B. *et al.* Disruption of the clock components CLOCK and BMAL1 leads to hypoinsulinaemia and diabetes. *Nature* **466**: 627–631.
69. Freathy, R.M. *et al.* Variants in *ADCY5* and near *CCN1* are associated with fetal growth and birth weight. *Nat. Genet.* **42**: 430–435.
70. Andersson, E.A. *et al.* Type 2 diabetes risk alleles near *ADCY5*, *CDKAL1* and *HHEX-IDE* are associated with reduced birthweight. *Diabetologia* **53**: 1908–1916.
71. Barker, D.J. *et al.* 1993. Type 2 (non-insulin-dependent) diabetes mellitus, hypertension and hyperlipidaemia (syndrome X): relation to reduced fetal growth. *Diabetologia* **36**: 62–67.
72. Pascoe, L. *et al.* 2007. Common variants of the novel type 2 diabetes genes *CDKAL1* and *HHEX/IDE* are associated with decreased pancreatic beta-cell function. *Diabetes* **56**: 3101–3104.
73. Palmer, N.D. *et al.* 2008. Quantitative trait analysis of type 2 diabetes susceptibility loci identified from whole genome association studies in the Insulin Resistance Atherosclerosis Family Study. *Diabetes* **57**: 1093–1100.
74. Staiger, H. *et al.* 2007. Polymorphisms within novel risk loci for type 2 diabetes determine beta-cell function. *PLoS One* **2**: e832.
75. Grarup, N. *et al.* 2007. Studies of association of variants near the *HHEX*, *CDKN2A/B*, and *IGF2BP2* genes with type 2 diabetes and impaired insulin release in 10,705 Danish subjects: validation and extension of genome-wide association studies. *Diabetes* **56**: 3105–3111.
76. Florez, J.C., 2008. Newly identified loci highlight beta cell dysfunction as a key cause of type 2 diabetes: where are the insulin resistance genes? *Diabetologia* **51**: 1100–1110.
77. Panhuysen, C.I. *et al.* 2003. A genome scan for loci linked to quantitative insulin traits in persons without diabetes: the Framingham Offspring Study. *Diabetologia* **46**: 579–587.
78. Bergman, R.N. *et al.* 2003. Minimal model-based insulin sensitivity has greater heritability and a different genetic basis than homeostasis model assessment or fasting insulin. *Diabetes* **52**: 2168–2174.
79. Araki, E. *et al.* 1994. Alternative pathway of insulin signalling in mice with targeted disruption of the IRS-1 gene. *Nature* **372**: 186–190.
80. Sparso, T. *et al.* 2008. The *GCKR* rs780094 polymorphism is associated with elevated fasting serum triacylglycerol, reduced fasting and OGTT-related insulinaemia, and reduced risk of type 2 diabetes. *Diabetologia* **51**: 70–75.
81. Orho-Melander, M. *et al.* 2008. 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* **57**: 3112–3121.

82. Maher, B. 2008. Personal genomes: the case of the missing heritability. *Nature* **456**: 18–21.
83. Florez, J.C., 2008. Clinical review: the genetics of type 2 diabetes: a realistic appraisal in 2008. *J. Clin. Endocrinol. Metab.* **93**: 4633–4642.
84. Via, M., C. Gignoux & E.G. Burchard. 2010. The 1000 Genomes Project: new opportunities for research and social challenges. *Genome. Med.* **2**: 3.
85. Thomas, D. 2010. Gene-environment-wide association studies: emerging approaches. *Nat. Rev. Genet.* **11**: 259–272.
86. Meigs, J.B. *et al.* 2008. Genotype score in addition to common risk factors for prediction of type 2 diabetes. *N. Engl. J. Med.* **359**: 2208–2219.
87. Lyssenko, V. *et al.* 2008. Clinical risk factors, DNA variants, and the development of type 2 diabetes. *N. Engl. J. Med.* **359**: 2220–2232.
88. de Miguel-Yanes, J.M. 2010. 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*. 2010 October 1 [Epub ahead of print].
89. Ramachandran, A. *et al.* 2006. The Indian Diabetes Prevention Programme shows that lifestyle modification and metformin prevent type 2 diabetes in Asian Indian subjects with impaired glucose tolerance (IDPP-1). *Diabetologia* **49**: 289–297.
90. Pan, X.R. *et al.* 1997. Effects of diet and exercise in preventing NIDDM in people with impaired glucose tolerance. The Da Qing IGT and Diabetes Study. *Diabetes Care* **20**: 537–544.
91. Tuomilehto, J. *et al.* 2001. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *N. Engl. J. Med.* **344**: 1343–1350.
92. Knowler, W.C. *et al.* 2002. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N. Engl. J. Med.* **346**: 393–403.
93. Florez, J.C. *et al.* 2006. *TCF7L2* polymorphisms and progression to diabetes in the Diabetes Prevention Program. *N. Engl. J. Med.* **355**: 241–250.
94. Hivert, M.F., A.S. Warner, P. Shrader *et al.* 2009. Diabetes risk perception and intention to adopt healthy lifestyles among primary care patients. *Diabetes Care* **32**: 1820–1822.
95. Grant, R.W. *et al.* 2009. The clinical application of genetic testing in type 2 diabetes: a patient and physician survey. *Diabetologia* **52**: 2299–2305.
96. Pearson, E.R. *et al.* 2006. Switching from insulin to oral sulfonylureas in patients with diabetes due to Kir6.2 mutations. *N. Engl. J. Med.* **355**: 467–477.
97. Sagen, J.V. *et al.* 2004. Permanent neonatal diabetes due to mutations in *KCNJ11* encoding Kir6.2: patient characteristics and initial response to sulfonylurea therapy. *Diabetes* **53**: 2713–2718.
98. Babenko, A.P. *et al.* 2006. Activating mutations in the *ABCC8* gene in neonatal diabetes mellitus. *N. Engl. J. Med.* **355**: 456–466.
99. Pearson, E.R. *et al.* 2003. Genetic cause of hyperglycaemia and response to treatment in diabetes. *Lancet* **362**: 1275–1281.
100. Sesti, G. *et al.* 2006. 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.* **91**: 2334–2339.
101. Feng, Y. *et al.* 2008. Ser1369Ala variant in sulfonylurea receptor gene *ABCC8* is associated with antidiabetic efficacy of gliclazide in Chinese type 2 diabetic patients. *Diabetes Care* **31**: 1939–1944.
102. Lehmann, J.M. *et al.* 1995. An antidiabetic thiazolidinedione is a high affinity ligand for peroxisome proliferator-activated receptor gamma (PPAR gamma). *J. Biol. Chem.* **270**: 12953–12956.
103. Kang, E.S. *et al.* 2005. Effects of Pro12Ala polymorphism of peroxisome proliferator-activated receptor  $\gamma$  2 gene on rosiglitazone response in type 2 diabetes. *Clin. Pharmacol. Ther.* **78**: 202–208.
104. Wolford, J.K. *et al.* 2005. Sequence variation in *PPARG* may underlie differential response to troglitazone. *Diabetes* **54**: 3319–3325.
105. Blucher, M., G. Lubben & R. Paschke. 2003. Analysis of the relationship between the Pro12Ala variant in the PPAR- $\gamma$ 2 gene and the response rate to therapy with pioglitazone in patients with type 2 diabetes. *Diabetes Care*. **26**: 825–831.
106. Snitker, S. *et al.* 2004. Changes in insulin sensitivity in response to troglitazone do not differ between subjects with and without the common, functional Pro12Ala peroxisome proliferator-activated receptor- $\gamma$ 2 gene variant: results from the Troglitazone in Prevention of Diabetes (TRIPOD) study. *Diabetes Care* **27**: 1365–1368.
107. Florez, J.C. *et al.* 2007. Effects of the type 2 diabetes-associated *PPARG* P12A polymorphism on progression to diabetes and response to troglitazone. *J. Clin. Endocrinol. Metab.* **92**: 1502–1509.
108. Pearson, E.R. 2009. Translating *TCF7L2*: from gene to function. *Diabetologia* **52**: 1227–1230.
109. Kimber, C.H. *et al.* 2007. *TCF7L2* in the Go-DARTS study: evidence for a gene dose effect on both diabetes susceptibility and control of glucose levels. *Diabetologia* **50**: 1186–1191.
110. Pearson, E.R. *et al.* 2007. Variation in *TCF7L2* influences therapeutic response to sulfonylureas: a GoDARTS study. *Diabetes* **56**: 2178–2182.
111. Winckler, W. *et al.* 2007. 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.
112. Gudmundsson, J., *et al.* 2007. Two variants on chromosome 17 confer prostate cancer risk, and the one in *TCF2* protects against type 2 diabetes. *Nat. Genet.* **39**: 977–983.
113. McCarthy, M.I. *et al.* 2008. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat. Rev. Genet.* **9**: 356–369.
114. Manolio, T.A. *et al.* 2009. Finding the missing heritability of complex diseases. *Nature* **461**: 747–753.