Candidate Gene Association Study in Type 2 Diabetes Indicates a Role for Genes Involved in β-Cell Function as Well as Insulin Action

Inês Barroso1†, Jian’an Luan2, Rita P. S. Middelberg2, Anne-Helen Harding2, Paul W. Franks2, Rupert W. Jakes2, David Clayton3, Alan J. Schafer1†, Stephen O’Rahilly4†, Nicholas J. Wareham2∗

1 Incyte, Palo Alto, California, United States of America, 2 Department of Public Health and Primary Care, University of Cambridge Institute of Public Health, Cambridge, United Kingdom, 3 Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research, Cambridge, United Kingdom, 4 Department of Clinical Biochemistry, University of Cambridge, Cambridge, United Kingdom

Type 2 diabetes is an increasingly common, serious metabolic disorder with a substantial inherited component. It is characterised by defects in both insulin secretion and action. Progress in identification of specific genetic variants predisposing to the disease has been limited. To complement ongoing positional cloning efforts, we have undertaken a large-scale candidate gene association study. We examined 152 SNPs in 71 candidate genes for association with diabetes status and related phenotypes in 2,134 Caucasians in a case-control study and an independent quantitative trait (QT) cohort in the United Kingdom. Polymorphisms in five of 15 genes (33%) encoding molecules known to primarily influence pancreatic β-cell function—ABCC8 (sulphonylurea receptor), KCNJ11 (Kir6.2), SLC2A2 (GLUT2), HNF4A (HNF4α), and INS (insulin)—significantly altered disease risk, and in three genes, the risk allele, haplotype, or both had a biologically consistent effect on a relevant physiological trait in the QT study. We examined 35 genes predicted to have their major influence on insulin action, and three (9%)—INSR, PIK3R1, and SOST—showed significant associations with diabetes. These results confirm the genetic complexity of Type 2 diabetes and provide evidence that common variants in genes influencing pancreatic β-cell function may make a significant contribution to the inherited component of this disease. This study additionally demonstrates that the systematic examination of panels of biological candidate genes in large, well-characterised populations can be an effective complement to positional cloning approaches. The absence of large single-gene effects and the detection of multiple small effects accentuate the need for the study of larger populations in order to reliably identify the size of effect we now expect for complex diseases.

Introduction

Type 2 diabetes is a serious metabolic disease associated with an increased risk of premature death and substantial disability, largely mediated through its adverse effects on the vasculature. The prevalence of the disease is increasing, and the World Health Organisation estimates suggest that by 2025 there will be 300 million affected individuals worldwide (King et al. 1998). The disorder is characterised by a combination of impaired insulin secretion and insulin action, both of which precede and predict the onset of disease (Weyer et al. 1999). Through its adverse impact on insulin action, obesity is a major risk factor for the disease. Although environmental factors, both post- and prenatal, play an important role in determining the risk of disease, a substantial body of evidence supports the notion that disease susceptibility is influenced by inherited factors (Zimmet 1982). While the molecular basis for several uncommon Mendelian forms of Type 2 diabetes have been defined (Vionnet et al. 1992; Yamagata et al. 1996a, 1996b; Horikawa et al. 1997; Stoffers et al. 1997; Barroso et al. 1999; Malecki et al. 1999; Savage et al. 2002), the nature and range of allelic variants conferring susceptibility to the more common forms of this disorder remain poorly defined. Many investigators have embarked on attempts to identify diabetes susceptibility genes through genome-wide linkage-based approaches using multigenerational pedigrees and/or large numbers of affected sibpairs. Regions of significant linkage, some of which have been replicated in more than one study, have been identified. To date, however, only Calpain 10 (CAPN10; LocusLink ID 11132) has emerged from such studies as a new putative diabetogene (Horikawa et al. 2000). While some subsequent studies have supported a role for the CAPN10 alleles originally described as susceptibility alleles, others have found associations with different alleles and some have found no association with this gene (Baier et al. 2000;
Cox 2001; Evans et al. 2001; Hegele et al. 2001; Tsai et al. 2001; Daimon et al. 2002; Elbein et al. 2002; Garant et al. 2002).

The positional cloning effort has been supplemented by a large number of studies examining specific candidate genes using a variety of methodologies, mostly of the case-control association design. Although many positive reports have emerged, few have been consistently replicated. Of these candidates, the most compelling evidence to date, generated from a meta-analysis of multiple published studies, is that a common amino acid variant in the N-terminus of the nuclear receptor peroxisome proliferator-activated receptor γ (PPARG; LocusLink ID 5468) confers significant protection against the development of Type 2 diabetes (Altshuler et al. 2000). More recently, evidence has accumulated supporting a role for the E23K variant of KCNJ11 (LocusLink ID 3767) in Type 2 diabetes predisposition (Hani et al. 1998; Glyn et al. 2001, 2003; Love-Gregory et al. 2003; Nielsen et al. 2003).

Whole-genome association studies in large case-control populations may ultimately have the greatest power to detect alleles of small but significant effects on the susceptibility to common diseases such as Type 2 diabetes. As yet, however, the resource implications of such an approach are prohibitive. In the meantime, knowledge of both mammalian biology and disease pathogenesis is progressing rapidly, and it is possible to identify a large panel of known genes, the dysfunction of which might reasonably be considered likely to contribute to Type 2 diabetes. In this study we have identified 152 informative single nucleotide polymorphisms (SNPs) in 71 such genes and, using these, have examined their association with Type 2 diabetes and related intermediate phenotypes in Caucasian subjects from the United Kingdom.

Results/Discussion

Overall Study Design

Candidate gene studies are based on selection of genes with a known or inferred biological function whose role makes it plausible that they may predispose to disease or the observed phenotype. These types of studies are similar to traditional epidemiological approaches in which an a priori hypothesis between exposure to a given factor (in this case, a genotype at a given locus) and disease is formulated. To date, most Type 2 diabetes candidate gene studies have largely lacked thoroughness and sensitivity, as they have tested a limited number of genes and variants in small populations or in populations that were poorly matched or phenotyped, frequently resulting in a lack of replication of the weak associations detected (Altshuler et al. 2000). Our strategy aimed to address these problems by a unique combination of features, including comprehensive SNP discovery in a large number of candidate genes, testing of a large number of SNPs, use of two independent populations, and analysis of haplotypes in addition to individual SNPs where possible.

Figure 1 illustrates the overall design of the study. On the basis of their known or putative role in glucose metabolism, 71 candidate genes were selected for study (Table 1). These were subdivided into three broad groups: (1) genes primarily involved in pancreatic β-cell function; (2) genes primarily influencing insulin action and glucose metabolism in the main target tissues, muscle, liver, and fat; and (3) other genes. This group includes genes that influence processes potentially relevant to diabetes, such as energy intake, energy expenditure, and lipid metabolism. A de novo search for common SNPs was undertaken using fluorescent single-stranded conformation polymorphism (SSSCP) examination of all coding regions and splice junctions in a variety of human populations. All genes were minimally screened against 47 samples of mixed ethnicity, providing 0.99 probability of detecting variants with a minor allele frequency of 0.05. Our ‘in-house’ polymorphism detection programme identified 954 SNPs in the 71 genes, with a range of allele frequencies from 0.003 to 0.50. Of the 152 SNPs chosen for further study (Table S1, found at http://dx.doi.org/10.1371/journal.pbio.0000020.s001), the great majority had a minor allele frequency of greater than 5%, but in a few instances less frequent variants were typed when the candidate gene had strong biological plausibility and there were no known polymorphisms of higher frequency at the time of SNP selection.

The 152 SNPs were genotyped in a population-based cohort of 517 unrelated Caucasians in the United Kingdom with Type 2 diabetes and an equal number of controls with normal glycated haemoglobin (HbA1c) levels, individually matched to cases by age, sex, and geographical location. A second independent population was also genotyped for the same 152 SNPs. This consisted of 1,100 middle-aged Caucasian subjects in the United Kingdom who had been extensively and serially phenotyped for glucose tolerance and insulin secretion, insulin action, and adiposity. In the first stage of data analysis, all SNPs (and haplotypes when multiple SNPs were present at the same gene) were examined for their association with diabetes in the case-control study using multiple models of inheritance. In the second phase of analysis, all SNPs and haplotypes showing statistically significant association with diabetes status in the first phase were examined for association with glucose levels and other intermediate phenotypes in the quantitative trait.
**Table 1. Genes with SNPs Genotyped in This Study**

| Group                        | Subgroup                          | Official Gene Symbol | Common Symbol | Number of SNPs Tested in This Study |
|------------------------------|-----------------------------------|----------------------|---------------|-------------------------------------|
| Pancreatic β-cell function   |                                   | CDX2                 | CDX3          | 1                                   |
|                              |                                   | NEUROD1              | 1             |
|                              |                                   | PAX4                 | 1             |
|                              |                                   | PAX6                 | 1             |
|                              |                                   | HNF4A                | 3             |
|                              |                                   | TCF1                 | HNF1A         | 1                                   |
|                              |                                   | TCF2                 | HNF1B         | 2                                   |
|                              |                                   | ABC2                 | SUR1          | 16                                  |
|                              |                                   | ADCYAP1R1            | PACAPR        | 1                                   |
|                              |                                   | CPE                  |               | 2                                   |
|                              |                                   | GCK                  | 1             |
|                              |                                   | GLP1R                | 5             |
|                              |                                   | INS                  | 1             |
|                              |                                   | KCNJ11               | KIR6.2        | 4                                   |
|                              |                                   | SLC2A2               | GLUT2         | 6                                   |
| Insulin action               | INSR signalling pathway           | AKT1                 | PKB           | 1                                   |
|                              |                                   | AKT2                 | 1             |
|                              |                                   | CAV3                 | 1             |
|                              |                                   | FOXO3A               | FKHRL1        | 1                                   |
|                              |                                   | FYN                  | 1             |
|                              |                                   | GRB10                | 1             |
|                              |                                   | GRB14                | 2             |
|                              |                                   | GRB2                 | 1             |
|                              |                                   | GRB7                 | 1             |
|                              |                                   | INSR                 | IR            | 7                                   |
|                              |                                   | IRS1                 | 2             |
|                              |                                   | PDE3B                | 1             |
|                              |                                   | PIK3CA               | 1             |
|                              |                                   | PIK3R1               | p85α          | 3                                   |
|                              |                                   | RPS6KA2              | 1             |
|                              |                                   | SLC2A4               | GLUT4         | 2                                   |
|                              |                                   | SOS1                 | 2             |
|                              |                                   | SOS2                 | 2             |
| Negative regulator of insulin action |                   | AHSG                 | 1             |
|                              |                                   | ENPP1                | PC-1          | 2                                   |
|                              |                                   | GH1                  | 1             |
|                              |                                   | INPPL1               | SHIP2         | 1                                   |
| Carbohydrate metabolism      |                                   | FBP1                 | 2             |
|                              |                                   | GPD1                 | 1             |
|                              |                                   | HK1                  | 2             |
|                              |                                   | HK2                  | 4             |
|                              |                                   | PFKM                 | 1             |
|                              |                                   | SLC2A5               | GLUT5         | 1                                   |
|                              |                                   | G6PT1                | 1             |
|                              |                                   | GYS1                 | 1             |
|                              |                                   | GYS2                 | 1             |
|                              |                                   | PPP1CC               | 1             |
|                              |                                   | PPP1R3A              | 3             |
|                              |                                   | PPP2R1A              | 1             |
|                              |                                   | PCK1                 | PEPCK         | 11                                  |
(QT) study population. The intermediate phenotypes chosen for study were fasting and 2-h post-glucose load plasma glucose levels (measures of glucose tolerance), fasting insulin (a measure of insulin sensitivity), 30-min insulin incremental response (a measure of β-cell function), and body mass index (BMI) (a measure of adiposity).

Power to detect an association is dependent on several factors: the frequency of the ‘predisposing’ allele, genotype, or haplotype; the accepted false-positive or Type 1 error rate ($\alpha$); and the odds ratio (OR) or effect size. Rarer alleles, genotypes, or haplotypes with small effects require larger sample sizes to attain the same power to detect an association, as compared to more frequent alleles or alleles with larger effects. At the time that we collected the populations and designed this study, our power calculations had shown that a sample size of 500 cases and 500 matched controls would have 80% power to detected effect sizes as small as 1.3–1.7 OR, depending on the frequency of the predisposing allele, with a 5% Type 1 error rate (Figure 2).

Overview of Results of Association Studies

Table S1 shows the genotype counts for all 152 SNPs in the case-control and QT studies. In the control subjects, 16 SNPs (10.6%) had a minor allele frequency below 5%; 19 (12.5%) had a minor allele frequency between 5% and 10%; and 117 (76.9%) had a minor allele frequency greater than or equal to 10%. Each variant was tested for association with disease status under several genetic models. Twenty SNPs in 11 different genes showed statistically significant association with disease status ($p < 0.05$) under at least one model (Table 2). The strongest statistical evidence for disease association was for genes SOS1 (LocusLink ID 6654), SLC2A2 (LocusLink ID 6514), PIK3R1 (LocusLink ID 5295), ABCC8 and KCNJ11 (LocusLink ID 6833), and INSR (LocusLink ID 3643). Of the 29 loci with multiple SNPs, only three—HNF4A (LocusLink ID

![Figure 2. Power Calculations](http://www.mc.vanderbilt.edu/prevmed/ps; DuPont and Plummer 1997).

DOI: 10.1371/journal.pbio.0000020.g002

Candidate genes, identified by official HUGO Gene Nomenclature Committee symbols, are grouped by known or putative biological function, with the number of genotyped polymorphisms per gene shown.

Table 1. Continued

| Group                        | Subgroup                        | Official Gene Symbol | Common Symbol | Number of SNPs Tested in This Study |
|------------------------------|---------------------------------|---------------------|---------------|------------------------------------|
| Others                       | Lipid metabolism               | CETP                | 3             |                                    |
|                              |                                 | FABP2               | 1             |                                    |
|                              |                                 | FABP4               | aP2           | 1                                  |
|                              |                                 | LIPC                | HL            | 3                                  |
|                              |                                 | LIPE                | HSL           | 1                                  |
|                              |                                 | PLCG1               | 1             |                                    |
|                              |                                 | PPARG               | 2             |                                    |
| Feeding behaviour/energy homeostasis | GAL                     | GALN               | 1             |                                    |
|                              |                                 | PPARGC1             | PG1           | 4                                  |
|                              |                                 | ADRB3               | 1             |                                    |
|                              |                                 | PYY                 | 1             |                                    |
| Others                       |                                 | USF1                | 1             |                                    |
|                              |                                 | KCNJ6               | KIR3.2        | 1                                  |
|                              |                                 | STXB3               | UNC-18c       | 1                                  |
|                              |                                 | SGNE1               | 1             |                                    |
|                              |                                 | NCOA1               | 1             |                                    |
|                              |                                 | RXRG                | 2             |                                    |
|                              |                                 | ABCC9               | SUR2          | 4                                  |
|                              |                                 | ENPP2               | PDNP2         | 2                                  |
|                              |                                 | GC                  | VDBP          | 4                                  |
|                              |                                 | NOS3                | eNOS          | 4                                  |

DOI: 10.1371/journal.pbio.0000020.t001

Candidate genes, identified by official HUGO Gene Nomenclature Committee symbols, are grouped by known or putative biological function, with the number of genotyped polymorphisms per gene shown.

DOI: 10.1371/journal.pbio.0000020.t001

(QT) study population. The intermediate phenotypes chosen for study were fasting and 2-h post-glucose load plasma glucose levels (measures of glucose tolerance), fasting insulin (a measure of insulin sensitivity), 30-min insulin incremental response (a measure of β-cell function), and body mass index (BMI) (a measure of adiposity).

Power to detect an association is dependent on several factors: the frequency of the ‘predisposing’ allele, genotype, or haplotype; the accepted false-positive or Type 1 error rate ($\alpha$); and the odds ratio (OR) or effect size. Rarer alleles, genotypes, or haplotypes with small effects require larger sample sizes to attain the same power to detect an association, as compared to more frequent alleles or alleles with larger effects. At the time that we collected the populations and designed this study, our power calculations had shown that a sample size of 500 cases and 500 matched controls would have 80% power to detected effect sizes as small as 1.3–1.7 OR, depending on the frequency of the predisposing allele, with a 5% Type 1 error rate (Figure 2).
Table 2. Genes with Variants Significantly Associated with Type 2 Diabetes Status

| Gene    | SNPID | SNP     | Allele 2 (22 + 12) OR p Value | Additive Effect OR p Value | Allele 2 (11 + 12) OR p Value |
|---------|-------|---------|-------------------------------|---------------------------|-------------------------------|
| SOS1    | 8     | IVS17+53| 0.58 0.0032                   | 0.58 0.0020               |                               |
| SLC2A2  | 21    | IVS5–15 | 1.44 0.0117                   | 1.35 0.0230               |                               |
|         | 23    | T198    | 1.39 0.0318                   | 1.36 0.0323               |                               |
|         | 24    | T110I   | 1.49 0.0059                   | 1.37 0.0114               |                               |
| PPARGC1 | 30    | T528    | 0.83 0.0378                   |                           | 0.67 0.0295                   |
|         | 31    | G482S   |                           |                           |                               |
| PIK3R1  | 42    | IVS4+82 | 1.41 0.0090                   | 1.34 0.0088               | 2.0 0.0258                    |
| INS     | 72    | 3p+9    |                           |                           |                               |
| KCNJ11  | 74    | 3p+215  | 0.75 0.0299                   | 0.76 0.0021               | 0.59 0.0027                    |
|         | 76    | A190    | 0.79 0.0127                   | 0.79 0.0127               | 0.62 0.0260                    |
|         | 77    | E23K    |                           |                           | 1.49 0.0033                    |
| ABCC8   | 79    | IVS38+54| 1.25 0.0131                   | 1.23 0.0256               | 1.68 0.0043                    |
|         | 81    | A13695  |                           |                           | 3.43 0.0163                    |
|         | 84    | IVS18–36|                           |                           | 3.90 0.0157                    |
|         | 87    | K649    |                           |                           |                               |
|         | 89    | IVS11–74|                           |                           | 2.82 0.0480                    |
| ABCC9   | 100   | IVS13–76|                           |                           | 1.99 0.0339                    |
| LIPC    | 114   | IVS1+49 | 0.76 0.0468                   | 0.77 0.0291               |                               |
| PYY     | 123   | IVS3+68 | 1.47 0.0240                   | 1.47 0.0157               |                               |
| INS     | 131   | IVS6+43 | 1.48 0.0039                   | 1.32 0.0119               |                               |

SNP identifiers (SNPID), OR, significance level (p value), and genetic model are shown. p values for the additive effect are for the test for a linear trend across the genotypes, which were coded as 0 = 11, 1 = 12, 2 = 22. Allele 2 dominant refers to a combination of 12 + 22 and allele 2 recessive refers to combination of 11 + 12.

DOI: 10.1371/journal.pbio.0000020.t002

3172) INS, and ABCC8–KCNJ11—showed significant association of particular haplotypes with disease status (Figure 3; Table 3). In only one case (HNF4A) was a haplotype significantly associated with disease risk (see below) when no significant association was seen with any individual SNP in that gene. Table S2, found at http://dx.doi.org/10.1371/journal.pbio.0000020.s002, shows the results of association studies undertaken in the QT population, further examining the SNPs that had shown significant association in the case-control study. Table 3 shows the relationship between disease-associated haplotypes at ABCC8–KCNJ11, HNF4A, and INS with intermediate phenotypes in the QT study. We now consider in more detail, the data for those genes where the strongest and most consistent effects were seen.

Genes Primarily Affecting β-Cell Function

ABCC8 and KCNJ11 (encoding, respectively, the sulphonylurea receptor and inwardly rectifying potassium channel KIR6.2). The genes encoding the two molecular components of the voltage-gated potassium channel of the pancreatic β-cell are located within 4.5 kb of each other on Chromosome 11. KCNJ11 encodes the channel protein KIR6.2 and ABCC8 encodes an ATP-binding cassette (ABC) transporter-containing transmembrane protein (SUR1) that is thought to regulate the activity of the channel and that also contains the site to which sulphonylurea antidiabetic drugs bind. Three SNPs in KCNJ11 were associated with disease under multiple genetic models. The strongest statistical association in this gene was with a 3′-SNP (SNP74; OR 0.59, p = 0.0027 under recessive model for allele 2) (see Table 2). In ABCC8, five SNPs were associated with disease status under multiple models; the strongest evidence for association with disease was with SNP79 and SNP81, respectively an intronic variant (OR 1.68, p = 0.0043; see Table 2) and a missense variant A1369S (OR 1.68, p = 0.0048; see Table 2). Although neither of these two SNPs was significantly associated with any trait in the QT study, two other SNPs showed effects in the QT study (see Table S2). SNP84 (IVS18–36) from ABCC8, which associated with increased disease risk (OR 3.43, p = 0.0163; see Table 2) also associated with increased BMI (mean 28.2 kg/m², 95% confidence interval [CI] [26.6, 29.9] for homozygous 11 and 5.27 mmol/l, 95% CI [5.19, 5.35] for homozygous 22 versus 5.30 mmol/l, 95% CI [5.26, 5.35] for heterozygous subjects; p = 0.016) and associated with borderline significance with higher fasting glucose (5.53 mmol/l, 95% CI [5.29, 5.77] for homozygous 22 versus 26.2 kg/m², 95% CI [25.9, 26.5] for homozygous 11 and 26.3 kg/m², 95% CI [25.7, 26.8] for heterozygous subjects; p = 0.016) and associated with borderline significance with decreased insulin secretion (23.6 pmol/mmol, 95% CI [23.6, 23.6]) under a recessive model for allele 2 (see Table S2). SNP87 (K649), which was also significantly associated with increased disease risk (OR 3.90, p = 0.0157; see Table 2), also showed borderline significant association with decreased insulin secretion (23.6 pmol/mmol, 95% CI [23.6, 23.6]) for homozygous 22 and 30.9 pmol/mmol/
mmol, 95% CI [28.6, 33.4] for heterozygous subjects; \( p = 0.054 \); see Table S2), consistent with a role for this gene in insulin secretion. Given the close physical proximity of \( ABCC8 \) and \( KCNJ11 \) and their role in the same functional unit, we performed haplotype reconstructions with data from both genes combined (see Figure 3A). Haplotype B was associated with increased disease risk (OR 1.46, 95% CI [1.14, 1.85]; data not shown), but did not show any significant association in the QT study (see Table 3).

Mutations in each of these genes have been associated with familial persistent hyperinsulinaemia hypoglycaemia of infancy (PHHI), a rare disorder of glucose homeostasis characterised by up-regulated insulin secretion despite severe hypoglycaemia. In addition, evidence for association of \( KCNJ11 \) DNA variants with Type 2 diabetes has been evaluated in multiple studies, and until recently these data have been conflicting. Several recent studies have, however, suggested a role for the aminoacid variant E23K in Type 2 diabetes susceptibility (Hani et al. 1998; Gloyn et al. 2001, 2003; Schwanstecher and Schwanstecher 2002; Love-Gregory et al. 2003; Nielsen et al. 2003). In total we tested four SNPs at the \( KCNJ11 \) locus for association with disease status; of these, three were tightly linked (data not shown) and all three had a statistically significant association with disease status (see Table 2). In our study we replicated the effect of the E23K polymorphism in Type 2 diabetes predisposition (KK homozygous, OR 1.49, \( p = 0.0333 \); see Table 2) with an OR estimate in agreement with that demonstrated by the meta-analysis of Nielsen et al. (2003); in addition, two other \( KCNJ11 \) SNPs associated with disease risk (SNP74 and SNP76). The recent evidence from multiple studies and from meta-analysis for association of the E23K SNP with Type 2 diabetes, along with in vitro studies using cell lines expressing the E23K mutation showing an increased stimulation threshold of insulin secretion (Schwanstecher et al. 2002), suggests that E23K is the functional variant leading to increased disease risk. Given our finding of high levels of linkage disequilibrium (LD) between SNP74 and SNP76 with E23K (data not shown),

Figure 3. Genes with Haplotypes Associated with Type 2 Diabetes
Genomic organization with exons (black boxes or vertical lines) and with genotyped SNPs and SNPs utilised in the haplotype reconstructions (in blue boxes) is shown. The most common haplotypes with population prevalence greater than 5% in the control population are shown, and the measure of LD (\( r^2 \)) is shown for a subset of the SNPs. (A) \( ABCC8–KCNJ11 \). (B) HNF4A. (C) INSR.
DOI: 10.1371/journal.pbio.0000020.g003
we adjusted the measures of association at these sites for the E23K genotype. These data suggest that these SNPs are independently associated with diabetes (Table 4).

ABCC8 variants have been associated with Type 2 diabetes in multiple studies (Inoue et al. 1996; T. Hansen et al. 1998; Hart et al. 1999b). However, a recent large study failed to replicate previous associations with Type 2 diabetes (Altshuler et al. 2000). In our study we found evidence for association with Type 2 diabetes in five of 16 ABCC8 SNPs tested. Owing to the physical mapping of ABCC8 in close proximity to

For case control, OR and 95% CI of haplotypes are shown. For QT, means and 95% CI of haplotypes are shown. Means were adjusted for age and sex, but not for BMI. Abbreviations: BMI, body mass index (kg/m^2); PG0, fasting plasma glucose (mmol/l); 2hPG, 2-h plasma glucose (mmol/l); INS0, fasting insulin (pmol/l); Ins inc, 30-min insulin increment (pmol/mmol). Associations significant at 0.10 and below are in italics bold, with association 0.05 or below in red bold.

Table 3. Association of ABCC8–KCNJ11, HNF4A, and INSR Haplotypes with Diabetes and QTs

| Gene   | Haplotype | p Value |
|--------|-----------|---------|
|        | A         | B       | C       |
|        | Diabetes  | BMI     | PG0     |
|        | 1.00      | 26.3    | 5.29    |
|        | 0.0846    | 0.26 (25.6, 26.4) | 5.26 (5.21, 5.31) |
|        | 0.78      | 1.07 (0.80, 1.44) | 5.30 (5.23, 5.37) |
|        | 0.004     | 0.94 (0.66, 1.34) | 5.24 (5.17, 5.32) |
|        | 0.359     | 26.0 (25.3, 26.7) | 5.88 (5.73, 6.09) |
|        | 0.454     | 5.89 (5.66, 6.12) | 5.91 (5.73, 6.09) |
|        | 0.081     | 5.92 (5.77, 6.08) | 5.93 (5.60, 5.89) |
|        | 0.072     | 5.89 (5.60, 5.89) | 5.91 (5.73, 6.09) |
|        | 0.578     | 40.9 (38.8, 43.0) | 42.8 (39.3, 46.6) |
|        | 0.028     | 30.9 (28.1, 34.0) | 31.5 (29.9, 33.3) |

Table 4. Association of KCNJ11 and ABCC8 Variants with Type 2 Diabetes Status Adjusted for E23K Genotype

| Gene | SNPID | SNP   | Allele 2 Dominant (22 + 12) | p Value | Allele 2 Recessive (11 + 12) | p Value |
|------|-------|-------|-----------------------------|---------|-----------------------------|---------|
|      |       |       | OR                          |         | OR                          |         |
|      |       |       | 0.78                        | 0.0846  | 0.78                        | 0.0256  |
|      |       |       | 0.0948                      | 0.82    | 0.0765                      | 0.67    |
|      |       |       | 0.3644                      | 2.09    | 0.0652                      | 3.67    |
|      |       |       | 0.5461                      | 1.68    | 0.1878                      | 2.85    |
|      |       |       | 0.9392                      | 1.05    | 0.7251                      | 3.23    |
|      |       |       | 0.5562                      | 1.16    | 0.2767                      | 4.36    |
|      |       |       | 0.1934                      | 0.91    | 0.4635                      | 3.16    |
|      |       |       | 0.0149                      |         | 0.0149                      |         |
|      |       |       | 0.0672                      |         | 0.0672                      |         |
|      |       |       | 0.0536                      |         | 0.0536                      |         |
|      |       |       | 0.1339                      |         | 0.1339                      |         |
|      |       |       | 0.0401                      |         | 0.0401                      |         |
|      |       |       | 0.0214                      |         | 0.0214                      |         |
|      |       |       | 0.0445                      |         | 0.0445                      |         |

For case control, OR and 95% CI of haplotypes are shown. For QT, means and 95% CI of haplotypes are shown. Means were adjusted for age and sex, but not for BMI. Abbreviations: BMI, body mass index (kg/m^2); PG0, fasting plasma glucose (mmol/l); 2hPG, 2-h plasma glucose (mmol/l); INS0, fasting insulin (pmol/l); Ins inc, 30-min insulin increment (pmol/mmol). Associations significant at 0.10 and below are in italics bold, with association 0.05 or below in red bold.

DOI: 10.1371/journal.pbio.0000200.k003

SNP identifiers (SNPID), OR, significance level (p value), and genetic model are shown. p Values for the additive effect are for the test for a linear trend across the genotypes, which were coded as 0 = 11, 1 = 12, 2 = 22. Allele 2 dominant refers to a combination of 12 + 22 and the allele 2 recessive refers to combination of 11 + 12.

DOI: 10.1371/journal.pbio.0000020.t004
all studies have been small and were insufficiently powered to diabetes, and most have reported negative results. However, increased disease risk, was associated with lower 2-h plasma glucose. No other significant associations with intermediate variables, such as increased diabetes risk were also all associated with lower 2-h plasma glucose, compared to the subjects with the other haplotypes (see Table 3). HNF4A maps to Chromosome 20 (Argyroskastritis et al. 1997) in a region that has been linked to Type 2 diabetes in multiple studies (Bowden et al. 1997; Ji et al. 1997; Zouali et al. 1997; Ghosh et al. 1999; Klupa et al. 2000; Permutt et al. 2001). This positional information, combined with the known role of major mutations at this gene in the causation of autosomal-dominant maturity-onset diabetes of the young (MODY), has led to HNF4A being considered as a strong candidate for involvement in Type 2 diabetes. However, most studies to date have failed to identify an association between variants at this locus and disease susceptibility (Moller et al. 1997; Malecki et al. 1998; Ghosh et al. 1999; Price et al. 2000). This study differs from all other previous reports in its examination of haplotypes, as well as in the fact that it included several SNPs not previously examined. Our findings lead us to speculate as to how a particular HNF4A haplotype might be associated with lower risk of diabetes and increased insulin secretory capacity. The fact that a multiplicity of heterozygous nonsense and missense mutations in HNF4A lead to an insulinopenic form of MODY strongly suggests that β-cell dysfunction is sensitive to the amount of HNF4A in the β-cell and that haploinsufficiency is the likely mode of molecular pathogenesis in that condition (Stoffel and Duncan 1997; Shih et al. 2000). It is plausible, therefore, that variants in this gene that enhance expression levels of the protein might lead to increased insulin secretory capacity and protection against diabetes.

INS (encoding insulin). The INS (LocusLink ID 3630) gene encodes the hormone preproinsulin, which upon propeptide cleavage generates mature insulin and C-peptide. We tested for association of a single SNP in the 3′-UTR (SNP72) of the insulin gene with disease status. This SNP was significantly associated with increased Type 2 diabetes risk under a recessive model for the T allele (OR 2.02, p = 0.0258) (see Table 2). In the QT study this SNP did not associate with any of the intermediate phenotypes studied. The insulin gene variable number tandem repeat (INS-VNTR) has been
extensively studied and is proposed to exert pleiotropic effects on birth weight and diabetes susceptibility (Huxtable et al. 2000). However, evidence for this has been conflicting and a role for INS in Type 2 diabetes predisposition has not been definitively established. The data for the single SNP we tested suggest that either the insulin gene or other loci in LD may be involved in Type 2 diabetes risk.

Genes Primarily Affecting Insulin Action

**INSR** (encoding the insulin receptor). At the INSR locus of the seven SNPs genotyped, we detected a single intronic SNP (SNP131) that was significantly associated with increased disease risk (OR 1.48, \(p = 0.0039\) for the dominant model for allele 2) (see Table 2). In the QT study, this SNP also had a nonsignificant association with increased 2-h glucose under a dominant model for allele 2 (see Table S2). Haplotype C (see Figure 3C) for INSR was associated with increased disease risk (1.34 mmol/l, 95% CI [1.05, 1.71]; data not shown); in the QT study, there was a nonsignificant trend for subjects carrying this haplotype to have increased values for fasting glucose (5.32 mmol/l, 95% CI [5.24, 5.39] versus 5.27 mmol/l, 95% CI [5.21, 5.33] for haplotype B and 5.27 mmol/l, 95% CI [5.23, 5.32] for haplotype A), 2-h glucose (6.00 mmol/l, 95% CI [5.80, 6.20] versus 5.78 mmol/l, 95% CI [5.61, 5.96] for haplotype B and 5.87 mmol/l, 95% CI [5.75, 5.99] for haplotype A), and fasting insulin (41.8 pmol/l, 95% CI [39.2, 44.5] versus 40.4 pmol/l, 95% CI [38.4, 42.5] for haplotype B and 41.3 pmol/l, 95% CI [39.6, 43.1] for haplotype A) (see Table 3). A role for INSR in Type 2 diabetes and related phenotypes has long been sought. Many studies initiated over the past decade have explored the possibility that DNA variants at this locus would not only cause rare syndromes of extreme insulin resistance, but would also associate with increased Type 2 diabetes risk. In particular, the role of the Val985Met in disease predisposition has been analysed in many different populations, but the data remain inconclusive, with some studies suggesting a role for this variant (Hart et al. 1996, 1999b), while others do not support this finding (O’Rahilly et al. 1991, 1992; L. Hansen et al. 1997). In this study we provide preliminary evidence for a role of INSR in diabetes susceptibility through genotyping of a previously untested SNP in case-control studies and via haplotype analysis using multiple SNPs in the gene.

**PIK3RI and SOS1.** The gene PIK3RI, encoding the p85z regulatory subunit of the phosphatidylinositol 3-kinase, is a logical candidate gene for involvement in the development of Type 2 diabetes owing to its role in insulin signal transduction. An intronic variant, SNP42, was associated with increased disease risk under two genetic models (OR 1.41, \(p = 0.0090\) for the allele 2 dominant and OR1.34, \(p = 0.0088\) for the additive model; see Table 2). In the QT study, SNP42 was significantly associated with increased BMI and showed a borderline significance with increased fasting insulin (measure of insulin resistance) under a dominant model for allele 2 (see Table S2). Obesity is a major risk factor for insulin resistance, and the observed increase in BMI coupled with increased insulin resistance in carriers of allele G at SNP42 suggests that variation at this gene may be increasing Type 2 diabetes risk through impaired insulin action. Other association studies at this locus have focused on investigating the Met326Ile variant in disease predisposition, with mostly negative results (T. Hansen et al. 1997; Kawanishi et al. 1997). One study did describe an association with disease status and with QTs underlying Type 2 diabetes (Baier et al. 1998). However, functional data for this polymorphism have suggested that the Ile326 variant may have only minor impact on signalling events (Baynes et al. 2000; Almind et al. 2002). Our data suggest that variation in this gene is a risk factor for the development of Type 2 diabetes, although further detailed studies will be required to elucidate the precise functional variants and mechanisms that lead to increased disease risk.

The gene **SOS1** (son of sevenless homolog 1 in *Drosophila*) encodes a guanine nucleotide exchange factor that functions in the transduction of signals that control cell growth and differentiation. We analysed two SNPs for association with disease status, a nonsynonymous SNP (N1011S) and an intronic variant (SNP8). While the nonsynonymous S1011 variant, which was very rare (minor allele, 0.003), did not associate with disease status, the intronic SNP was highly significantly associated with decreased disease risk (OR 0.58, \(p = 0.0032\)) (see Table 2), despite not showing any effects in the QT study. To our knowledge, this is the first investigation into the role of **SOS1** in Type 2 diabetes risk. The data presented here suggest that further investigation into the potential role of common variants at this gene and diabetes risk is warranted.

Other Genes

**PPARGC1** (LocusLink ID 10891) encodes a transcriptional coactivator of nuclear receptors with a critical role in regulating multiple aspects of energy metabolism, including adaptive thermogenesis (Puigserver et al. 1998), mitochondrial biogenesis (Wu et al. 1999), fatty acid β-oxidation (Vega et al. 2000), the control of hepatic gluconeogenesis (Herzig et al. 2001; Yoon et al. 2001), and the control of glucose uptake (Michael et al. 2001). **PPARGC1** SNP30 (Thr528Thr), which was associated with decreased disease risk (see Table 2), was rather surprisingly associated with decreased insulin secretion in the QT study (see Table S2). In this locus, Thr528Thr has not been previously associated with diabetes, and our data most likely reflect stochastic variation at this site. The Gly482Ser has in some studies been shown to be associated with increased Type 2 diabetes risk (Ek et al. 2001), but not in others (Hara et al. 2002; Lacquemant et al. 2002; Muller et al. 2003), and has additionally been associated with insulin resistance (Hara et al. 2002), obesity indices in women (Estherbauer et al. 2002), and mean insulin secretory response and lipid oxidation (Muller et al. 2003). In our study, this allele was not associated with increased diabetes risk, but rather was associated with a lower risk of diabetes under a recessive model (OR 0.67, \(p = 0.0295\)) (see Table 2). The opposing results for this polymorphism and the fact that the amino acid change Gly482Ser is unlikely to be a major functional change (Estherbauer et al. 2002) may indicate that the contributing functional polymorphism may be an unidentified variant in LD with the Gly482Ser.

Amongst the remaining genes tested, of particular interest are the results observed in **PYY** (encoding polypeptide YY; LocusLink ID 5697). An intronic variant, IVS3+68, showed a significant association with increased Type 2 diabetes risk under two genetic models (OR 1.47, \(p = 0.0240\) in the allele 2 dominant; OR 1.47, \(p = 0.0157\) in the additive effect allele 2) (see Table 2), but no evidence of association with underlying
traits was observed in the QT study. Early functional studies suggested an inhibitory role of PYY in glucose-stimulated insulin secretion (Bertrand et al. 1992; Nieuwenhuizen et al. 1994), which led us to evaluate the potential role of variants at this gene in Type 2 diabetes predisposition. Recent data have shown that the peptide PYY<sub>3-36</sub> encoded by this gene inhibits food intake and reduces weight gain when injected in rats, while physiological infusions of PYY<sub>3-36</sub> in humans decreased food intake by 33% (Batterham et al. 2002). Although our data do not show an association between the intronic variant SNP122 with BMI, they suggest a putative role for PYY in Type 2 diabetes predisposition. As we only tested a non-coding variant in PYY, it is possible that the association is due to other contributing variants within the gene and that a link between those and BMI is still plausible.

In the genes ABCC9 (LocusLink ID 10060) and LIPC (LocusLink ID 3990), single SNPs of modest significance were associated with disease status in the case control and therefore are not discussed further here.

**Examination of Other Previously Reported Associations**

We were unable to confirm some associations observed in other studies. The PPARG Pro12Ala Pro allele has previously been shown to confer susceptibility to Type 2 diabetes, with the Ala allele providing a decreased risk (Altshuler et al. 2000). Our results for this polymorphism show the same direction and magnitude of effect for Ala/Ala versus Ala/Pro and Pro/Pro genotypes (OR 0.54; derived from data in Table 2), but the association was not statistically significant (p = 0.2269). The lower limit of the 95% CI for the protective effect of the Ala allele (OR 1.08, 95% CI [0.82,1.42], p = 0.583) is still consistent with the results of the meta-analysis by Altshuler et al. (2000). Our study was not sufficiently powered to detect the small effects expected for the predisposing Pro allele. The study only had 25.4% power to detect an OR of 1.25 for the Pro allele that occurred in 89.4% of our control population. It is also possible that the Pro12Ala variant does not affect diabetes susceptibility in this population, because of the dependence of the allele effect on environmental factors such as dietary fat composition (Luan et al. 2001).

In the ENPP1 (LocusLink ID 5167) gene (commonly known as PG-1) the K121Q polymorphism has variably been found to be both associated with increased Type 2 diabetes risk (Gu et al. 2000; Rasmussen et al. 2000; Hegele et al. 2001) and with insulin resistance QTs (Pizzuti et al. 1999; Gu et al. 2000; Rasmussen et al. 2000). In our study we did not find evidence for association between the K121Q polymorphism and Type 2 diabetes (OR 1.10, p = 0.5277 for the dominant model for allele 2; OR 1.07, p = 0.6290 for the additive effect for allele 2; OR 0.86, p = 0.7496 for a recessive model for allele 2; data derived from Table 2). Analysis of this allele in our QT study showed that QQ individuals have higher mean BMI levels compared to carriers of the K121 allele (28.3 kg/m<sup>2</sup> [26.4, 30.2] versus 26.1 [25.6, 26.7] in KQ subjects, 26.3 [25.9, 26.6] in KK subjects; data not shown). PPP1R3A (protein phosphatase 1 regulatory subunit 3; LocusLink ID 5506), which encodes the muscle-specific regulatory subunit of PP1, has been investigated as a potential diabetogene. Evidence for a role of the PPP1R3A D905Y polymorphism in Type 2 diabetes risk has also been conflicting (L. Hansen et al. 1995, 2000; Hegele et al. 1998; Xia et al. 1998). While in this study we did not find an association between the D905Y variant and disease risk, we have previously described an association between a rare frameshift and premature stop variant with Type 2 diabetes risk under a dominant model (OR 5.03, p = 0.0110) in this population (Savage et al. 2002).

**Concluding Remarks**

This study, which to our knowledge is the largest of its kind yet reported in Type 2 diabetes, has provided evidence for the existence of variants in certain key candidate genes that influence the risk of Type 2 diabetes and, in some cases, has afforded clues as to the pathophysiological mechanism whereby those effects on disease risk might be mediated. By its very nature, any study of candidate genes, however large, is restricted in scope, and it is likely that other variants (namely in regulatory regions, which we did not cover) in the genes that we have considered, as well as ones that we have not, may exist and have effects equal to or greater than those we have demonstrated. In addition, this study is not intended to be an ‘exclusion study,’ as many issues that relate to coverage of any given gene, environmental risk factors, and power in our populations do not allow us to definitively assert that negative findings correspond to genes that truly do not play a role in Type 2 diabetes predisposition.

The power of our study to detect small effects in uncommon variants was low. Evidence from many recent studies now suggests that in Type 2 diabetes the effects are likely in the range of OR 1.15-1.5. It is clear that much larger studies than that reported here are required for such effects (Figure 4), in particular when adjusting to a lower Type 1 error rate of 0.01% to compensate for multiple testing. The significance of the associations we report have been described without adjustment for the number of tests undertaken, and thus the group of positive associations is likely to contain a proportion that is falsely positive. There is no consensus about the ideal method for adjusting the probability of an observation occurring by chance for multiple testing. The simple Bonferroni correction would constitute overadjustment because the 152 genetic markers in this study are not independent. In addition, in the false-discovery rate method (Benjamini and Hochberg 1995), it is assumed that all N tests are carried out simultaneously, which may not correspond to reality if groups genotype one set of SNPs, as in this study, but then report results for additional SNPs at a later date. It is not clear whether the number of tests N should reflect the number to date or the number one might potentially undertake by continuing working through projects like these. An alternative Bayesian approach leading to a ‘genome-wide’ significance level for association, such as has been done for whole-genome linkage studies (Lander and Kruglyak 1995), might be preferable. However, this also runs into difficulties. In studies that are not based on fine-mapping of linkage intervals, but rather on candidate genes selected on the basis of data from other studies, including previous reports of association, it is unclear what level of prior probability of association should be used. As a result of this uncertainty about the appropriate method of correction for multiple testing, our preferred strategy is to report the number of tests done and to encourage readers to interpret the significance tests in that light, acknowledging that the results will require replication in other cohorts.

Although we have not undertaken a formal replication of
have a substantial environmental influence modulated by the causation of Type 2 diabetes, it is tempting to speculate would be premature to put forward any definitive model for diabetes risk through impairment of insulin action. While it all of which are thought to have their major influence on function among those that were found to be associated with genetic architecture of Type 2 diabetes may vary between variants in their own particular populations. Indeed, as the identification of pathophysiological mechanisms by which the association bias. The associations we describe are highly biologically plausible and in many of the genes are supported by associations with multiple SNPs at the same locus. These include genes affecting both insulin secretion and insulin action. Given the importance of both insulin resistance and defective β-cell function to the pathogenesis of Type 2 diabetes, it is intriguing that we have found a disproportionate representation of genes affecting pancreatic β-cell function among those that were found to be associated with diabetes risk. This contrasts with the impact of known environmental factors and their correlates (e.g., high-fat diet, lower physical activity, obesity, and central fat distribution), all of which are thought to have their major influence on diabetes risk through impairment of insulin action. While it would be premature to put forward any definitive model for the causation of Type 2 diabetes, it is tempting to speculate that the ‘insulin resistance’ component of the disease may have a substantial environmental influence modulated by polygenic effects, some of which may relate to molecules identified in this and other studies. On the other hand, the ability of the pancreatic β-cell to continue to secrete sufficient insulin to maintain life-long normoglycaemia may be more profoundly influenced by genetic factors, some of which are reported herein. It will be critical to examine the functional consequences of such variants, a task that will be particularly challenging when it comes to genes influencing human β-cell function, as it is entirely possible that this disproportionate representation of β-cell genes may be a reflection of our success in choosing diabetes genes in each of the candidate genes in the major groupings.

The success of the approach presented here, although necessarily limited in scope, suggests that the systematic examination of panels of biological candidate genes in large, well-characterised populations may usefully complement positional approaches to the identification of allelic variants conferring susceptibility to complex polygenic disease. The detection of multiple small gene effects accentuates the need for larger populations in order to reliably identify the types of effects (OR 1.15–1.5) we now expect for complex diseases.

**Materials and Methods**

**Methods for SNP discovery and SNP selection for genotyping.** SNP discovery was performed by a high-SSCP-based analysis, as previously described (Thorpe et al. 1999). Genomic structure was determined for all genes, and primers were designed to span the exons and splice junctions. To detect common variants, genes were screened against one or more of a variety of different DNA panels, which included a 47-member multiethnic human diversity panel (comprised of 17 Europid, seven Hispanic, 13 East Asian, and ten African-American subjects), our 129-member severe insulin-resistant cohort (Rosso et al. 1999), a panel of 47 Euro-African samples, a panel of 47 African-American samples, and in some cases a panel of 94 samples (half European and half Asian Indian). Some genes had only partially screened coding sequence and splice junctions at the time of SNP selection for genotyping. In addition, we had access to an internal database of in silico SNPs that had been validated against 100 samples. Choice of polymorphisms for further testing in association studies was not constrained by the type of variant (e.g., nonsynonymous, silent, noncoding), although higher priority was given to variants with a likely effect on protein structure and function. Polymorphisms with a minor allele frequency greater than or equal to 5% were selected for further testing in population-based studies. In some instances, polymorphisms of lower allele frequency were genotyped to examine whether lower frequency variants with high penetrance might account for some cases of polygenic disease. Polymorphisms of lower frequency were also genotyped when, at the time of selection for genotyping, no other variants of known frequency were identified in the gene to test.

**Populations for SNP genotyping.** The Cambridgeshire Type 2 Diabetes Genes (POULTON et al. 2002; Hallsall et al. 2003) consists of a collection of 517 Type 2 diabetics and 517 matched controls. The samples were a random sample of Europid men and women with Type 2 diabetes, aged 47–75 years, from a population-based diabetes register in a geographically defined region in Cambridgeshire, United Kingdom. The presence of Type 2 diabetes in these subjects was defined as onset of diabetes after the age of 30 years without use of insulin therapy in the first year after diagnosis. The control subjects were individually age-, gender-, and geographical location-matched to each of the cases. Controls were not matched by BMI to cases. Potential controls that had HbA1c levels greater than 6% were excluded, as this group may contain a higher proportion of individuals with previously undiagnosed diabetes. HbA1c was assessed using high performance liquid chromatography on a BioRad Diamat 33 (Hercules, California, United States), according to the method of Standing and Taylor (1992). The coefficient of variation (CV) was 5.6% at the lower end of the range (mean = 4.94%) and 3.0% at the upper end (mean = 9.76%). Further details on the characteristics of the subjects are shown in Table 5.

The QT study population is a collection of 1,100 samples collected for the Ely Study, a prospective population-based cohort study of the aetiology and pathogenesis of Type 2 diabetes and related metabolic...
Table 5. Study Subjects in the Cambridgeshire Case-Control Study

|                  | Men |          |          |          | Women |          |          |          |
|------------------|-----|----------|----------|----------|-------|----------|----------|----------|
|                  | Case | Control | p Value  |         | Case  | Control | p Value  |         |
|                  | (n = 331) | (n = 331) |          |          | (n = 186) | (n = 186) |          |          |
| Age (years)      | 63.7 (0.4) | 63.7 (0.4) | 0.96     |          | 63.3 (0.6) | 63.5 (0.6) | 0.93     |          |
| Age at diagnosis | 55.8 (0.5) |          | —        |          | 55.5 (0.7) |          | —        |          |
| BMI (kg/m²)      | 28.6 (0.2) | 27.1 (0.2) | <0.0001  |          | 31.6 (0.5) | 27.6 (0.4) | <0.0001  |          |
| Number of individuals in treatment for diabetes | 92 (diet alone) | 45 (diet alone) 137 |          |          |          |          |          |          |
| (tablets or/and insulin) |      |          |          |          |      |          |          |          |

Data are means and standard error is in parentheses. DOI: 10.1371/journal.pbio.0000020.t005

Methods for genotyping. Genotyping was performed using an adaptation of the fluorescence polarization template-directed incorporation (FP-TDI) method described by Chen et al. (1999). In short, PEP-amplified DNA samples were PCR-amplified in 8 μl reactions with primers flanking the variant site; unincorporated dNTP and remaining unused primer were degraded by exonuclease I and shrimp alkaline phosphatase at 37°C for 45 min before the enzymes were heat-inactivated at 95°C for 15 min. At the end of the reaction, the samples were held at 4°C. Single base primer extension reactions were performed as previously described (Chen et al., 1999), and allele detection was performed by measuring fluorescence polarization on an LJL Analyst fluorescent reader (Molecular Devices, Sunnyvale, California, United States). The PEP protocol was specifically developed and tested to ensure that allele bias was not introduced during the amplification process. A minimum of 12% internal replicate samples within each population (case control and QT) were included in all genotyping tests to assess genotyping accuracy. Only assays that provided 100% concordance between replicates were analyzed for association. The genotyping pass rate was greater than 90% once a working assay had been established. There was an 85% success rate for an SNP to be converted into a working assay at the first attempt, with a number of failed assays recovered by designing an assay to the reverse strand.

Statistical analysis. All analyses used SAS 8.02 (SAS Institute, Cary, North Carolina, United States) or Stata 7.0 (Stata Corporation, College Station, Texas, United States) statistical programs, unless otherwise stated. Agreement with Hardy–Weinberg equilibrium was tested using a $X^2$ goodness-of-fit test. The disequilibrium coefficient for the controls ($r^2$) was calculated (Weinberg 1964).

For the case-control study, tests for association with disease status under dominant, additive, and recessive models were undertaken using univariate logistic regression analysis. Dominance was defined in terms of allele 2 effects; in the dominant allele 2 model, homozygous subjects for allele 1 were compared with carriers of allele 2; in the recessive allele 2 model, carriers of allele 1 were compared with homozygous subjects for allele 2. In some cases, a large number of polymorphisms within a gene were typed. To reduce complexity, a subset of markers within a gene associated with diabetes status was identified using backward logistic regression. Any polymorphism that had a $p$ value greater than 0.1 was removed from the model. The genotypes were assumed as having additive effects. $p$ values for the additive effect are for the test for a linear trend across the genotypes, which were coded as 0 = 11, 1 = 12, 2 = 22. Where the subset consisted of more than one polymorphism within a gene, haplotype analysis was performed. To avoid possible errors due to either genotyping or the estimation process, only haplotypes that had a frequency greater than 5% were considered for further analysis. Haplotype frequencies were estimated using maximum-likelihood methods. A log-linear model embedded with the expectation-maximization algorithm was fitted to a frequency table (Chiano and Clayton 1998; Mander 2001). Differences in haplotype distributions between the diabetic and non-diabetic groups were examined using a likelihood-ratio statistic (Mander 2001). To obtain separate ORs for each haplotype, the most common haplotype was used as the reference category. CIs were obtained using a profile-likelihood approach (Mander 2001).

For the QT study, the distributions of fasting plasma glucose, 2-h plasma glucose, fasting plasma insulin, and insulin increment were skewed and were thus normalised by logarithmic transformation. Baseline and follow-up measurements of BMI, fasting and 2-h plasma glucose, fasting plasma insulin, and 30-min insulin increment during an oral glucose tolerance test were collected. Where two measures were available, the mean was used. Otherwise, a single measure (either baseline or follow-up) was used for further analysis. The subset of SNPs identified in the case-control study was used. In separate dominant, additive, and recessive models, adjusting for age and sex, genotype differences in these measurements were modelled using the General Linear Model procedure in the statistical package SAS. For each individual, a list of possible haplotypes and their probabilities were obtained using Snphap software (http://www-gene.cimr.cam.ac.uk/clayton/software). Haplotypes with a frequency greater than 5% were the same as those reconstructed in the case-control study. Only haplotypes that had a frequency greater than 5% and individuals that had at least one marker typed were considered for analysis. As currently haplotype analysis software cannot handle repeated measurements, the average of two measurements was used for further analysis. Associations of haplotypes (adjusted for age and sex) with the QTs were determined by cluster-weighted regression analysis, thereby taking into account non-independent multiple observations from an individual (Huber 1967; White 1980, 1982). QT means and their 95% CI were estimated for each haplotype.

Supporting Information

Table S1. Genotype Counts and Frequencies for All SNPs Genotyped in This Study

View online at DOI: 10.1371/journal.pbio.0000020.s001 (79 KB XLS).

Table S2. Single SNP Associations with QTs

View online at DOI: 10.1371/journal.pbio.0000020.s002 (174 KB DOC).
Acknowledgments

NJW is a Wellcome Trust Senior Clinical Fellow. The Ely Study is supported by the Medical Research Council, SO is supported by the Wellcome Trust and the Juvenile Diabetes Research Foundation. This study drew upon the combined efforts of many individuals at Incyte Genomics Cambridge, to whom we extend our grateful appreciation. We also thank Suzannah Bumpstead, Bill Bottomley, and Stephan Collins for technical assistance. We are grateful to anonymous reviewers for helpful comments and suggestions.

Conflicts of Interest. Portions of the research were supported by Incyte Corporation (formerly Incyte Genomics). This collaboration provided access to certain technologies and scientific expertise for basic research in the genetics of diabetes in return for possible shared intellectual property. The work described here is not subject to any patent filings or intellectual property protection and restrictions from use of the information. IB and AJS were employees of Incyte at the time the research took place.

Author Contributions. IB, AJS, SO, and NW conceived and designed the experiments. IB, AJS, and NW performed the experiments or the fieldwork in the studies contributing data. IB, JL, RPSM, A-HH, PWF, RWJ, and NW analysed the data. DC contributed analysis tools and advice. All authors wrote the paper.

References

Almind K, Delahaye L, Hansen T, van Obberghen E, Pedersen O, et al. (2002) Characterization of the Met326Val variant of phosphatididylinositol 3-kinase p85α. Proc Natl Acad Sci U S A 99: 2124–2128.

Alpha B, Cox L, Crowther N, Clark PM, Hales CN (1992) Sensitive amplified immunoenzymometric assays (IEMA) for human insulin and intact proinsulin. Eur J Clin Chem Clin Biochem 30: 27–32.

Ahlskog D, Hirschorn JN, Klammernik M, Lindsey CM, Voh MC, et al. (2000) The common PPARG Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. Nat Genet 26: 76–80.

Argyrokastaris A, Kamaki S, Kapsetaki M, Kritis A, Talianidis I, et al. (1997) Human hepatocyte nuclear factor-4 gene maps to 2q12-q13.1 between PLCG1 and D20S17. Hum Genet 99: 233–236.

Baier LJ, Wiedrich C, Hanson RL, Bogardus C (1998) Variant in the regulatory subunit of phosphatidylinositol 3-kinase (p85α). Preliminary evidence indicates a potential role of this variant in the acute insulin response and type 2 diabetes in Pima women. Diabetes 47: 973–975.

Baier LJ, Permana PA, Yang X, Pratley RE, Hanson RL, el al. (2000) A calpain-10 gene polymorphism is associated with reduced muscle mRNA levels and insulin resistance. J Clin Invest 106: R69–R75.

Baronti M, Alcolado JC, Pozzi A, Cavallo MG, Li SR, et al. (2003) Polymorphisms at the GLUT2 (beta-cell/liver) glucose transporter gene and noninsulin-dependent diabetes mellitus (NIDDM): Analysis in affected pedigree members. Clin Genet 41: 229–324.

Barroso I, Gurnell M, Crowley VE, Agostini M, Schwabe JW, et al. (1999) Dominant negative mutations in human PPARG associated with severe insulin resistance, diabetes mellitus and hypertension. Nature 402: 880–883.

Batterham RL, Cowley MA, Small CJ, Herzog H, Cohen MA, et al. (2002) Gut hormone PYY(3–36) physiologically inhibits food intake. Nature 418: 650–654.

Baynes KC, Beeton CA, Panayotou G, Stein R, Soos M, et al. (2000) Natural immunoassay (IFA) for human insulin and intact proinsulin. Eur J Clin Chem Clin Biochem 38: 210–212.

Bertrand G, Gross R, Roye M, Ahren B, Ribes G (1992) Evidence for a direct inhibitory effect of PYY on insulin secretion in rats. Pancreas 7: 595–600.

Besselsen DB, Gross R, Roye M, Ahren B, Ribes G (1992) Evidence for a direct inhibitory effect of PYY on insulin secretion in rats. Pancreas 7: 595–600.

Beynon J, Andersen G, Urhammer SA, Gaede PH, Drivsholm T, et al. (2001) Genetic and environmental associations with glucose tolerance and the missing data problem. Ann Hum Genet 62 (Pt 1): 55–60.

Bihari G, Neely J, ללכת לפני השון, ומכים מאסטרים של מחקרים אחרים. מככרים让我们说的, סקר do's עיתונים, ונתונים do'.area hecho análisis instrumentos de la medicina. Om J Hum Genet 4: 1351–1350.

Bihari G, Neely J, ללכת לפני השון, ומכים מאסטרים של מחקרים אחרים. מככרים让我们说的, סקר do's עיתונים, ונתונים do'area hecho análisis instrumentos de la medicina. Om J Hum Genet 4: 1351–1350.

Bihari G, Neely J, ללכת לפני השון, ומכים מאסטרים של מחקרים אחרים. מככרים让我们说的, סקר do's עיתונים, ונתונים do'area hecho análisis instrumentos de la medicina. Om J Hum Genet 4: 1351–1350.

Bihari G, Neely J, ללכת לפני השון, ומכים מאסטרים של מחקרים אחרים. מככרים让我们说的, סקר do's עיתונים, ונתונים do'area hecho análisis instrumentos de la medicina. Om J Hum Genet 4: 1351–1350.

Bihari G, Neely J, ללכת לפני השון, ומכים מאסטרים של מחקרים אחרים. מככרים让我们说的, סקר do's עיתונים, ונתונים do'area hecho análisis instrumentos de la medicina. Om J Hum Genet 4: 1351–1350.

Bihari G, Neely J, ללכת לפני השון, ומכים מאסטרים של מחקרים אחרים. מככרים让我们说的, סקר do's עיתונים, ונתונים do'area hecho análisis instrumentos de la medicina. Om J Hum Genet 4: 1351–1350.
Hart LM, Stoilk RP, Dekker JM, Nijpels G, Grobbee DE et al. (1999b) Prevalence of variants in candidate genes for type 2 diabetes mellitus in The Netherlands. The Rotterdam study and the Hoorn study. J Clin Endocrinol Metab 84: 1002–1006.

Hegele RA, Harris SB, Zimman B, Wang J, Cao H, et al. (1998) Variation in the AU(AT)-rich element within the 3′-untranslated region of PFP1-R3 is associated with variation in plasma glucose in aboriginal Canadians. J Clin Endocrinol Metab 83: 3980–3983.

Hegele RA, Harris SB, Zimman B, Hanley AJ, Cao H (2001) Absence of association of two diabetes 2 genes with CAPN10 and PC-1 polymorphisms in Oji Cree. Diabetes Care 24: 1498–1499.

Herzig S, Long F, Jalal US, Hedrich S, Quinn R, et al. (2001) CREB regulates hepatic gluconeogenesis through the coactivator PGC-1. Nature 413: 179–183.

Horikawa Y, Iwasaki N, Hara M, Furuta H, Hinokio Y, et al. (1997) Mutation in the hepatocyte nuclear factor-1 beta gene (TCF2) associated with MODY. Nat Genet 17: 384–385.

Horikawa Y, Oda N, Cox NJ, Li X, Orho-Melander M, et al. (2000) Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. Nat Genet 26: 163–175.

Huher PJ (1967) The behavior of maximum likelihood estimates under nonstandard conditions. In: Proceedings of the fifth Berkeley symposium on mathematical statistics and probability. Berkeley: University of California Press. 1: 221–253.

Huxtable SJ, Saker PJ, Haddad L, Walker M, Frayling TM, et al. (2000) Analysis of cancer-susceptibility loci provides evidence for linkage and association between the insulin gene and type 2 diabetes mediated exclusively through paternally transmitted class III variable number tandem repeat alleles. Diabetes 49: 126–130.

Inoue H, Ferrer J, Welling CM, Elbein SC, Hoffman M, et al. (1996) Sequence variations in the suflurbox enzyme receptor (SUR) gene are associated with NIDDM in Caucasians. Diabetes 45: 825–831.

Ji L, Malecki M, Warram JH, Yang Y, Rich SS, et al. (1997) New susceptibility locus for NIDDM is localized to human chromosome 2q. Diabetes 46: 876–881.

Kawashima M, Tamori Y, Masugi J, Mori H, Ito C, et al. (1997) Prevalence of a polymorphism of the phosphofructokinase 3 kinase p85 alpha regulatory subunit (cald 326 Met→Ile) in Japanese NIDDM patients. Diabetes Care 20: 1045.

King H, Aubert RE, Herman WH (1998) Global burden of diabetes, 1995–2025: Prevalence, numerical estimates, and projections. Diabetes Care 21: 1414–1431.

Klupa T, Malecki MT, Pezzolesi M, Ji L, Curtis S, et al. (2000) Further evidence for a susceptibility locus for type 2 diabetes on chromosome 2q. Diabetes 49: 2212–2216.

Kunst A, Draeger B, Ziegenhorn J (1983) UV-methods with hexokinase and glucose-6-phosphate dehydrogenase. In: Bergmeyer H, editor. Methods of enzymatic analysis. Weinheim: Verlag Chemie. pp. 163–172.

Lacquemant C, Chikri M, Boutin P, Samson C, Froguel P (2002) No association for gene-nutrient interaction at the PPARalpha locus. Diabetes 50: 686–689.

Lander E, Kruglyak L (1995) Genetic dissection of complex traits: Guidelines for model development. Science 266: 20–23.

Lewontin RC (1964) The interaction of selection and linkage. I. General considerations. Genetics 49: 31–56.

Li SR, Alcolado JC, Stocks J, Baroni MG, Oelbaum RS, et al. (1991) Genetic studies of genetic variability of the glucose transporter 2 promoter in Caucasians. Diabetes 40: 980–983.

Lewittes JC, Gheorghiade M, Cobb JP, Crosson JT, Nissen SE, et al. (2001) Contribution of pro- and anti-inflammatory cytokines to the pathogenesis of dilated cardiomyopathy. Circulation 103: 151–155.

Lewittes JC, Gheorghiade M, Cobb JP, Crosson JT, Nissen SE, et al. (2001) Contribution of pro- and anti-inflammatory cytokines to the pathogenesis of dilated cardiomyopathy. Circulation 103: 151–155.

Li SR, Alcolado JC, Stocks J, Baroni MG, Oelbaum RS, et al. (1991) Genetic studies of genetic variability of the glucose transporter 2 promoter in Caucasians. Diabetes 40: 980–983.

Lewittes JC, Gheorghiade M, Cobb JP, Crosson JT, Nissen SE, et al. (2001) Contribution of pro- and anti-inflammatory cytokines to the pathogenesis of dilated cardiomyopathy. Circulation 103: 151–155.

Lewittes JC, Gheorghiade M, Cobb JP, Crosson JT, Nissen SE, et al. (2001) Contribution of pro- and anti-inflammatory cytokines to the pathogenesis of dilated cardiomyopathy. Circulation 103: 151–155.
White HA (1980) Heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48: 817–830.

White HA (1982) Maximum likelihood estimation of misspecified models. Econometrica 50: 1–25.

Wu Z, Puigserver P, Anderson U, Zhang C, Adelmant G, et al. (1999) Mechanisms controlling mitochondrial biogenesis and respiration through the thermogenic coactivator PGC-1. Cell 98: 115–124.

Xia J, Scherer SW, Cohen PT, Majer M, Xi T, et al. (1998) A common variant in PPP1R3 associated with insulin resistance and type 2 diabetes. Diabetes 47: 1519–1524.

Yamagata K, Furuta H, Oda N, Kaisaki PJ, Menzel S, et al. (1996a) Mutations in the hepatocyte nuclear factor-4α gene in maturity-onset diabetes of the young (MODY1). Nature 384: 458–460.

Yamagata K, Oda N, Kaisaki PJ, Menzel S, Furuta H, et al. (1996b) Mutations in the hepatocyte nuclear factor-4α gene in maturity-onset diabetes of the young (MODY1). Nature 384: 455–458.

Yoon JC, Puigserver P, Chen G, Donovan J, Wu Z, et al. (2001) Control of hepatic gluconeogenesis through the transcriptional coactivator PGC-1. Nature 413: 131–138.

Zimmet P (1982) Type 2 (noninsulin-dependent) diabetes: An epidemiological overview. Diabetologia 22: 399–411.

Zouali H, Hani EH, Philippi A, Vionnet N, Beckmann JS, et al. (1997) A susceptibility locus for early-onset noninsulin dependent (type 2) diabetes mellitus maps to chromosome 20q, proximal to the phosphoenolpyruvate carboxykinase gene. Hum Mol Genet 6: 1491–1498.