ELECTRONIC SUPPLEMENTARY MATERIAL

STUDY SAMPLES AND TYPE 2 DIABETES DIAGNOSIS

Stage 1: discovery samples

Chinese University of Hong Kong

We have previously described the study design, ascertainment, inclusion criteria and phenotyping procedures of participants included in this study [1]. All participants were of southern Han Chinese ancestry residing in Hong Kong. Type 2 diabetes was diagnosed according to the 1998 World Health Organization (WHO) criteria. Patients with classic type 1 diabetes with acute ketotic presentation or continuous requirement of insulin within 1 year of diagnosis were excluded. Written informed consent was obtained from all participants. This study was approved by the Clinical Research Ethics Committee of the Chinese University of Hong Kong.

In the first stage discovery cohort (stage 1), we performed genome-wide scanning in 202 Hong Kong Chinese individuals (102 type 2 diabetes patients and 100 healthy controls) (Hong Kong GWAS 1 cohort). We selected 102 type 2 diabetes cases with young-onset diabetes diagnosed at age <40 years, positive family history and overweight, and 100 controls using the criteria of 1) no past diagnostic history of type 2 diabetes, impaired fasting glucose (IFG) or impaired glucose tolerance (IGT); 2) without family history of type 2 diabetes; and 3) with BMI ≤ 25 kg/m^2 and waist circumference ≤ 90 cm and 80 cm for men and women, respectively.

In addition, we genome-scanned 1,068 Hong Kong Chinese individuals (400 type 2 diabetes patients from the Hong Kong Diabetes Registry and 668 non-diabetic controls) (Hong Kong GWAS 2 cohort). The 668 diseased controls were individuals aged ≥ 16 years old with diseases other than type 2 diabetes that included 457 epilepsy cases, 111 eczema cases and 100 healthy individuals without hypertension

(recruited from the control arm of a hypertension study).

Shanghai Jiao Tong University Diabetes Study (SJTUDS)

All samples were recruited from Shanghai Diabetes Institute of Shanghai Jiao Tong University. The genome-wide scan was performed in 394 samples, including 197 type 2 diabetes patients and 197 normal glucose regulation controls. The type 2 diabetes patients were probands of diabetic pedigrees with fasting plasma glucose ≥ 7.0 mmol/L and/or 2-h post plasma glucose ≥ 11.1 mmol/L who were diagnosed before 40 years old. Type 1 diabetes and mitochondrial diabetes were excluded based on clinical, immunological and genetic criteria. The controls were individuals with normal glucose regulation with fasting plasma glucose < 6.1 mmol/L and 2-h plasma glucose < 7.8 mmol/L as assessed by standard 75g OGTTs, negative diabetic family history, aged over 50 years old and with a BMI below 23kg/m^2.

Clinical studies

All Hong Kong and Shanghai Chinese individuals underwent detailed clinical investigation. For the Hong Kong study, fasting blood samples of all studied participants were collected for fasting plasma glucose (FPG) and fasting plasma insulin (FPI). For the Shanghai study, fasting blood samples were obtained from cases, whilst controls had blood collected at baseline, and 120 min during a 75g oral glucose tolerance test (OGTT). Detailed clinical information, including age at diagnosis and presence of diabetic complications, were documented in all cases as described [35]. Homeostasis model assessment of insulin resistance (HOMA-IR) was calculated as $(FPI \times FPG) \div 22.5$, and homeostasis model assessment of beta-cell function (HOMA-β) was calculated as $FPI \times 20 \div (FPG-3.5)$ [2]. Stumvoll indices for beta-cell function were calculated for Shanghai controls which underwent OGTT with measurement of

insulin levels [3].

Stage 2: de novo replication samples

Hong Kong replication 1 (case-control cohort)

The case cohort consisted of 5,366 unrelated type 2 diabetes patients (mean age 56.7 ± 13.4 years, 45.1% male, mean duration of T2D 6.6 ± 6.9 years) selected from the Hong Kong Diabetes Registry (HKDR). The control cohort consisted of 2474 individuals ascertained from 3 sources: a) 985 adolescents from a community-based school survey of cardiovascular risk factors (mean age 15.5 ± 1.9 years, 44.2% male), [4] b) 513 hospital staff and adult volunteers participating in a community-based health screening program (mean age 42.0 ± 10.4 years, 47% male) [4] and c) 976 healthy elderly individuals free of diabetes selected from 4,000 elderly individuals recruited from the community (mean age 72.3 ± 5.3 years, 51.4% male) [5].

Shanghai replication 1 (case-control cohort)

All participants were of Chinese Han ancestry and resided in Shanghai and the nearby area. The type 2 diabetes patients (n=4,036) were selected from the Shanghai Diabetes Institute Inpatient Database (SHDIID), which recruited participants from inpatients in the Department of Endocrinology and Metabolism, Shanghai Jiao Tong University Affiliated Sixth People's Hospital beginning in 2001. Type 1 diabetes and mitochondrial diabetes were excluded based on clinical, immunological and genetic criteria. The controls (n=3,964) were selected from Shanghai Diabetes Study (SHDS) I and II, which are community-based surveys of diabetes performed in 1998–2001 (SHDS I) and 2007–2008 (SHDS II). The controls had fasting plasma glucose <6.1 mmol/L and 2-h plasma glucose < 7.8 mmol/L as assessed by standard 75g OGTTs, and had no family history of diabetes mellitus.

Hong Kong replication 2 (family cohort)

The study design, ascertainment, inclusion criteria and phenotyping of the Hong Kong Family Diabetes Study (HKFDS) have been described elsewhere [6]. Briefly, 325 individuals with type 2 diabetes (mean age 48.0 ± 14.4, 40.6% male) and 368 control subjects (mean age 37.0 ± 13.6, 41.0% male) were selected from 178 families consisting of siblings, parents, spouses, and offspring (> 16 years) ascertained through a proband with type 2 diabetes. Patients with clinical or autoimmune type 1 diabetes and families with known maturity-onset diabetes of the young or mitochondrial DNA nucleotide 3243 A>G mutations were excluded.

Shanghai replication 2 (family cohort)

We recruited 248 type 2 diabetes pedigrees with 657 type 2 diabetes patients and 168 individuals with normal glucose regulation. Pedigrees with any type 1 diabetic patient or mitochondria diabetic patient were excluded.

Stages 3 and 4: in silico replication samples

BioBank Japan (BBJ)

The individuals were recruited from several medical institutes in Japan, including Fukujuji Hospital, Iizuka Hospital, Iwate Medical University School of Medicine, National Hospital Organization Osaka National Hospital, Nihon University, Nippon Medical School, Osaka Medical Center for Cancer and Cardiovascular Diseases, The Cancer Institute Hospital of Japanese Foundation for Cancer Research, Tokushukai Hospitals and Tokyo Metropolitan Geriatric Hospital. We selected type 2 diabetes cases from individuals registered as having type 2 diabetes. Diabetes was originally diagnosed according to the World Health Organization (WHO) criteria. type

2 diabetes was clinically defined as disease with a gradual adult onset. Individuals who tested positive for antibodies to glutamic acid decarboxylase (GAD) and those diagnosed with a mitochondrial disease or MODY were not included in the case group. Controls were individuals registered as individuals not having type 2 diabetes but with diseases other than type 2 diabetes, comprised of 13 distinct diseases, or healthy volunteers. Individuals who had been analyzed in the previous report [7] were excluded from the present study. Altogether, 4,878 individuals with type 2 diabetes (case 1, age, 65.8 ±10.0 years; BMI, 24.1 ± 3.8 kg/m2; (all values are expressed as mean ± s.d.)) and 3,345controls (control 1, age, 52.5 ± 15.2 years; BMI, 22.5 ± 3.8 kg/m2; (all values are expressed as mean ± s.d.)) were genotyped. A total of 7,541 individuals belonging to the Hondo cluster (4,470 cases and 3,071 controls) were selected. We directly genotyped samples using Illumina HumanHap610-Quad (type 2 diabetes patients) and 550K BeadChip (controls).

Korea Association Resource Study (KARE)

Two KARE study cohorts were established as part of the Korean Genome Epidemiology Study (KoGES) in 2001 [8]. The sampling base for both cohorts was in the Kyung Gi-Do province, close to Seoul, the capital of the Republic of Korea. Both cohorts were designed to allow longitudinal prospective study and adopted the same investigational strategy. Participants have been examined every two years since baseline (2001). More than 260 traits have been extensively examined through epidemiological surveys, physical examinations, and laboratory tests applied to cohort members. A total of 10,038 individuals from KARE study cohorts were genotyped with Affymetrix Genome-Wide Human SNP array 5.0. to undertake a large-scale GWA analysis for type 2 diabetes and numerous complex quantitative traits. Of them, 1,042 individuals were included as type 2 diabetes cases according to the following

criteria: (1) treatment of type 2 diabetes, (2) fasting plasma glucose ≥ 7 mmol/L or plasma glucose 2-h after ingestion of 75gm oral glucose load ≥ 11.1 mmol/L and (3) age of disease onset ≥ 40 years. The inclusion criteria of nondiabetic control individuals (n = 2,943) were as follows: (1) no history of diabetes and (2) fasting plasma glucose < 5.6 mmol/L and plasma glucose 2-h after ingestion of 75gm oral glucose load < 7.8 mmol/L at both baseline and follow up studies.

Singapore Chinese, Malaysian and Indian populations (case-control cohorts)

The Singapore case-control study contained individuals from three sources: 1) 1998 Singapore National Health Survey (NHS98); 2) Singapore Malay Eye Study (SiMES); and 3) Singapore Diabetes Cohorts Study (SDCS) [9]. In the NHS98 cohort, individuals with fasting plasma glucose (FPG) <6.0 mmol/L and 2 hour post-challenge glucose (2HPG) <7.0 mmol/L were defined as NGT. Individuals with FPG ≥6.0 and <7.0 mmol/L, and 2HPG ≥ 7.0 and <7.8 mmol/L, were defined as having IFG. Individuals with FPG ≥7.0 mmol/L, and 2HPG ≥7.8 and <11.1 mmol/L, were defined as IGT. A total of 838 IFG/IGT individuals were excluded, leaving 3,032 NGT controls (2,196 Chinese, 472 Malays, and 364 Indians) available for selection. Individuals from the NHS98 and SDCS cohorts with: 1) a reported history of type 2 diabetes; 2) FPG ≥7.0 mmol/L; or 3) 2HPG≥11.1 mmol/L were defined as cases. 453 NHS98 cases (224 Chinese, 113 Malays, and 116 Indians) and 1,703 SDCS cases (1,317 Chinese, 256 Malays, and 130 Indians) were available for selection. In the SiMES cohort, individuals with non-fasting PG< 11.1 mmol/L and HbA_{1c} <6.1% (2 SD above the mean for the nondiabetic population) were defined as controls (N=1,785). Individuals with a reported history of type 2 diabetes or non-fasting PG level ≥11.1 mmol/L were defined as cases (N =707). From these three sources, we included: 1) 2,010 type 2 diabetes cases and 1,945 NGT controls of Chinese ancestry,

2) 794 type 2 diabetes cases and 1,240 NGT controls of Malaysian ancestry, 3) and 977 type 2 diabetes cases and 1,169 NGT controls of Indian ancestry, for analysis.

Chinese Hans (case-control cohort)

This study included 1999 type 2 diabetic cases and 1976 nondiabetic controls drawn from the Nutrition and Health of Aging Population in China (NHAPC) study (312 cases and 815 controls), the Gut Microbiota and Obesity Study (GTOS) (82 cases and 163 controls), the Fudan-Huashan Study (FHS) (807 cases and 339 controls), and the Beijing Diabetes Survey (798 cases and 659 controls). Details on the study have been published previously [10]. All participants were unrelated Chinese Hans from Beijing and Shanghai. Type 2 diabetic cases were identified as those with previously diagnosed type 2 diabetes and current use of antidiabetic treatment or who meet the following criteria: 1) 30≤ age ≤70, 2) fasting plasma glucose ≥7.0 mmol/l, 3) 2-h postprandial plasma glucose ≥11.1 mmol/l in a standard 75 g oral glucose tolerance test (OGTT) or plasma HbA1c ≥6.5%. The nondiabetic controls were selected according to the following criteria: 1) age ≥30, 2) no past history of diagnosis of diabetes and no family history of diabetes, 3) fasting glucose <5.6 mmol/l, 4) 2-h OGTT <7.8 mmol/l and/or HbA1c content <5.6%. The studies were approved by local ethnic committees of each participating institution, and written informed consents were obtained from all participants.

The DNA samples were genotyped using the Illumina Human660W-Quad BeadChip (Illumina, Inc., San Diego, CA, USA), and the genotypes were called using the Illumina GenCall algorithm. Some of samples were excluded if their genotype call rates < 97%, excessive heterozygosity, gender mismatches between the reported and genetically inferred gender or duplicates among other samples. Principle component

analysis was used to assess population structure of the samples and detected outliers along the first two eigenvectors which were excluded from further analyses. SNPs with genotype call rate < 95%, MAF < 0.5% or deviation from Hardy-Weinberg equilibrium ($p < 10^{-6}$) in control groups were also excluded. After all these quality control processes, 495,686 SNPs and 3,712 samples, including 1,873 type 2 diabetic cases (861 male and 1012 female) and 1,839 controls (803 male and 1036 female), remained for association analyses.

DIAGRAM Consortium (case-control cohort)

DIAGRAM+ study comprised 8,130 type 2 diabetes cases and 38,987 controls from eight type 2 diabetes GWAS of European descent, including the Wellcome Trust Case Control Consortium (WTCCC), Diabetes Genetics Initiative (DGI) and Finland-US Investigation of NIDDM genetics (FUSION) scans (the individuals of a previous joint analysis), with those from scans performed by deCODE genetics, the Diabetes Gene Discovery Group, the Cooperative Health Research in the Region of Augsburg group (KORAgen), the Rotterdam study and the European Special Population Research Network (EUROSPAN) (for details of sample characteristics, please see Supplementary Table 2 in reference [11]).

THE MUTHER (Multiple Tissue Human Expression Resource) CONSORTIUM

The MuTHER resource (www.muther.ac.uk) includes LCLs, skin and adipose tissue derived simultaneously from a subset of well-phenotyped healthy female twins from the TwinsUK adult registry. Whole-genome expression profiling of the samples, each with either two or three technical replicates, were performed using the Illumina Human HT-12 V3 BeadChips (Illumina Inc) according to the protocol supplied by the manufacturer. Log2 transformed expression signals were normalized separately per

tissue as follows: quantile normalization was performed across technical replicates of each individual followed by quantile normalization across all individuals. Genotyping was done with a combination of Illumina arrays (HumanHap300, HumanHap610Q, 1M□Duo and 1.2MDuo 1M. Untyped HapMap2 SNPs were imputed using the IMPUTE software package (v2). The number of adipose samples with genotypes and expression values is 776. Association between all SNPs (MAF>5%, IMPUTE info >0.8) within a gene or within 1MB of the gene transcription start or end site and normalized expression values were performed with the GenABEL/ProbABEL packages using the polygenic linear model incorporating a kinship matrix in GenABEL followed by the ProbABEL mmscore score test with imputed genotypes. Age and experimental batch were included as cofactors.

Additional information on methods

Imputation

Before imputation, the SNP ID (rs number) was standardized according to dbSNP build 129, and their physical positions were standardized according to build 36. SNPs were further excluded sequentially if: 1) their polymorphisms were A/T or C/G; 2) absent from dbSNP build 129; 3) genotyped in only case or only control cohorts; 4) absent from the 1000 Genomes reference panel for CHB+JPT (March 2010 release of pilot project 1). For each sample set in stage 1, all SNPs were aligned to the positive strand and imputed (via the MLE approach) using the MACH 1.0 software [12]. We imputed genotypes for autosomal SNPs that were present in the March 2010 release of phased 1000 Genomes genotype data from 60 CHB+JPT founders [13], but were not present in the genome-wide chip or did not pass direct genotyping QC. Cases and controls were merged into a single cohort for imputation based on 440,194, 435,953 and 274,752 quality autosomal SNPs in Hong Kong GWAS 1, Hong Kong GWAS 2 and Shanghai GWAS case-control cohorts, respectively. For the Hong Kong GWAS 1 cohort, one-step imputation was applied. For the Hong Kong GWAS 2 and Shanghai GWAS cohorts, two-step imputation was used to improve imputation efficiency, by randomly selecting 100 cases and 100 controls for model parameter estimation first before imputation.

A total of 3,356,999 SNPs in Hong Kong GWAS 1, 3,355,668 SNPs in Hong Kong GWAS 2 and 3,087,246 SNPs in Shanghai GWAS that passed the quality control filters (with predicted Rsq≥0.5 and and MAF≥0.1) after imputation were included in the association analyses for individual cohorts. Rsq is an imputation accuracy measure to estimate the squared correlation between the true and imputed

genotypes.

Genomic control

Genomic control (GC) was applied to correct for relatedness of the individuals and adjust for potential population stratification [14]. The inflation factor λ was estimated by taking the median of the distribution of the χ^2 statistic from all quality SNPs in association test, and then divide by the median of the expected χ^2 distribution. We calculated the *p* values corrected for genomic control by dividing the observed χ^2 statistic by λ. In this study, we adjusted for GC in two levels. Firstly, we corrected each individual study for λ separately in directly genotyped and imputed SNPs. Then we further adjusted for GC on the meta-analysis results.

eQTL analysis

Gene expression analysis was performed using data available from the MuTHER (Multiple Tissue Human Expression Resource) Consortium and GenCord Projects. The MuTHER resource (www.muther.ac.uk) includes lymphoblastoid cell lines (LCLs), skin and adipose tissue derived simultaneously from a subset of well-phenotyped healthy female twins from the TwinsUK adult registry [15]. Whole-genome expression profiling of the samples, each with either two or three technical replicates, were performed using the Illumina Human HT-12 V3 BeadChips (Illumina Inc) according to the protocol supplied by the manufacturer. Log_2 transformed expression signals were normalized separately per tissue as follows: quantile normalization was performed across technical replicates of each individual followed by quantile normalization across all individuals. Genotyping was done with a combination of Illumina arrays (HumanHap300, HumanHap610Q, 1MDuo and 1.2MDuo). Untyped HapMap2 SNPs

were imputed using the IMPUTE software package (v2). The number of adipose samples with genotypes and expression values is per tissue was 778 for LCLs, 667 in skin and 776 in adipose. Association between rs10229583 (MAF > 5%, IMPUTE info > 0.8) and the normalized mRNA expression values of genes within 1MB of the SNP were performed with the GenABEL/ProbABEL packages using the polygenic linear model incorporating a kinship matrix in GenABEL followed by the ProbABEL mmscore test with imputed genotypes. Age and experimental batch were included as cofactors. A multiple-testing correction was applied to the *cis*-association results. Genome-wide FDR of 1% for multiple testing corresponds to a *p* value threshold of $p=5.1\times10^{-5}$ in adipose tissue, 7.8×10^{-5} in LCLs and 3.81×10^{-5} in skin.

Additional eQTL analysis was carried out by identifying all SNPs in strong LD with rs10229583 using data extracted from SNAP (http://www.broadinstitute.org/mpg/snap/ldsearch.php). Selection criteria were as follows: (1) 1000 Genome Pilot 1; (2) r^2 limit: None; (3) Population Panel: CHBJPT; and (4) Distance Maximum 500kb. Twenty-four tagSNPs in strong LD to rs10229583 with $r^2 \geq 0.8$ were extracted. The maximum genomic distance to rs10229583 was 34kb. The LD between these tagSNPs and the MuTHER eQTL peaks within the dataset were then examined.

The GenCord project is the study of association analysis for eQTL to nearby SNPs in three cell types (primary fibroblasts, lymphoblastoid cells and T cells) from the umbilical cords of 75 individuals [16]. We used the Java based application Genevar (http://www.sanger.ac.uk/resources/software/genevar/) [17] to retrieve the relevant association results of our variant and expression of genes within 1MB of the SNP from the GenCord project. We also examined the LD between the variant and the eQTL peak within the dataset.

Gene network analysis

Experimentally validated gene-gene interactions (pathways, genetic interactions and physical interactions) were obtained from GeneMania [18]. Data was updated by GeneMania as of Feb 2012. The weight on each edge, as represented by the thickness of the edge, was computed by GeneMania and reflects the degree of confidence of the relationships between any gene pair within the network. Genes that are identified from the MuTHER eQTL analysis were entered as prior knowledge, and were used to guide the gene-network building. Pathway information was obtained from Pathway Commons [19] and genetic interactions were obtained from BioGrid [20].

References

[1] Ng MC, Park KS, Oh B, et al. (2008) Implication of genetic variants near TCF7L2, SLC30A8, HHEX, CDKAL1, CDKN2A/B, IGF2BP2, and FTO in type 2 diabetes and obesity in 6,719 Asians. Diabetes 57: 2226-2233

[2] Matthews RD, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC (1985) Homeostasis model assessment: insulin resistance and B cell function from fasting plasma glucose and insulin concentrations in man. Diabetologia 28: 412-419

[3] Stumvoll M, Mitrakou A, Pimenta W, et al. (2000) Use of the oral glucose tolerance test to assess insulin release and insulin sensitivity. Diabetes Care 23: 295-301

[4] Ng MC, Tam CH, So WY, et al. (2010) Implication of genetic variants near NEGR1, SEC16B, TMEM18, ETV5/DGKG, GNPDA2, LIN7C/BDNF, MTCH2, BCDIN3D/FAIM2, SH2B1, FTO, MC4R, and KCTD15 with obesity and type 2 diabetes in 7705 Chinese. J Clin Endocrinol Metab 95: 2418-2425

[5] Tang NL, Liao CD, Ching JK, et al. (2010) Sex-specific effect of Pirin gene on bone mineral density in a cohort of 4000 Chinese. Bone 46: 543-550

[6] Ng MC, So WY, Cox NJ, et al. (2004) Genome-wide scan for type 2 diabetes loci in Hong Kong Chinese and confirmation of a susceptibility locus on chromosome 1q21-q25. Diabetes 53: 1609-1613

[7] Yamauchi T, Hara K, Maeda S, et al. (2010) A genome-wide association study in the Japanese population identifies susceptibility loci for type 2 diabetes at UBE2E2 and C2CD4A-C2CD4B. Nat Genet

[8] Cho YS, Go MJ, Kim YJ, et al. (2009) A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. Nat Genet 41: 527-534

[9] Tan JT, Ng DP, Nurbaya S, et al. (2010) Polymorphisms identified through genome-wide association studies and their associations with type 2 diabetes in Chinese, Malays, and Asian-Indians in Singapore. J Clin Endocrinol Metab 95: 390-397

[10] Li H, Gan W, Lu L, et al. A Genome-Wide Association Study Identifies GRK5 and RASGRP1 as Type 2 Diabetes Loci in Chinese Hans. Diabetes

[11] Voight BF, Scott LJ, Steinthorsdottir V, et al. (2010) Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet 42: 579-589

[12] Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR (2010) MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol 34: 816-834

[13] A map of human genome variation from population-scale sequencing. Nature 467:

1061-1073
[14]Devlin B, Roeder K (1999) Genomic control for association studies. Biometrics 55: 997-1004
[15]Nica AC, Parts L, Glass D, et al. (2011) The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. PLoS Genet 7: e1002003
[16]Dimas AS, Deutsch S, Stranger BE, et al. (2009) Common regulatory variation impacts gene expression in a cell type-dependent manner. Science 325: 1246-1250
[17]Yang TP, Beazley C, Montgomery SB, et al. (2010) Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. Bioinformatics 26: 2474-2476
[18]Warde-Farley D, Donaldson SL, Comes O, et al. (2010) The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res 38: W214-220
[19]Cerami EG, Gross BE, Demir E, et al. (2011) Pathway Commons, a web resource for biological pathway data. Nucleic Acids Res 39: D685-690
[20]Stark C, Breitkreutz BJ, Chatr-Aryamontri A, et al. (2011) The BioGRID Interaction Database: 2011 update. Nucleic Acids Res 39: D698-704