ARTICLE

Genome-wide association study in a Chinese population identifies a susceptibility locus for type 2 diabetes at 7q32 near *PAX4*

R. C. W. Ma · C. Hu · C. H. Tam · R. Zhang · P. Kwan · T. F. Leung · G. N. Thomas ·
M. J. Go · K. Hara · X. Sim · J. S. K. Ho · C. Wang · H. Li · L. Lu · Y. Wang · J. W. Li ·
Y. Wang · V. K. L. Lam · J. Wang · W. Yu · Y. J. Kim · D. P. Ng · H. Fujita ·
K. Panoutsopoulou · A. G. Day-Williams · H. M. Lee · A. C. W. Ng · Y-J. Fang ·
A. P. S. Kong · F. Jiang · X. Ma · X. Hou · S. Tang · J. Lu · T. Yamauchi · S. K. W. Tsui ·
J. Woo · P. C. Leung · X. Zhang · N. L. S. Tang · H. Y. Sy · J. Liu · T. Y. Wong ·
J. Y. Lee · S. Maeda · G. Xu · S. S. Cherny · T. F. Chan · M. C. Y. Ng · K. Xiang ·
A. P. Morris · DIAGRAM Consortium · S. Keildson · The MuTHER Consortium ·
R. Hu · L. Ji · X. Lin · Y. S. Cho · T. Kadowaki · E. S. Tai · E. Zeggini · M. I. McCarthy ·
K. L. Hon · L. Baum · B. Tomlinson · W. Y. So · Y. Bao · J. C. N. Chan · W. Jia

Abstract

Aims/hypothesis Most genetic variants identified for type 2 diabetes have been discovered in European populations. We performed genome-wide association studies (GWAS) in a Chinese population with the aim of identifying novel variants for type 2 diabetes in Asians.

R. C. W. Ma, C. Hu, C. H. Tam and R. Zhang contributed equally to this study.

J. C. N. Chan and W. Jia jointly directed this study.

A full list of members of the DIAGRAM and MuTHER Consortia is provided in the electronic supplementary material.

Electronic supplementary material The online version of this article (doi:10.1007/s00125-013-2874-4) contains peer-reviewed but unedited supplementary material, which is available to authorised users.

R. C. W. Ma (✉) · C. H. Tam · P. Kwan · J. S. K. Ho · Y. Wang ·
V. K. L. Lam · H. M. Lee · A. C. W. Ng · A. P. S. Kong · J. Woo ·
G. Xu · B. Tomlinson · W. Y. So · J. C. N. Chan
Department of Medicine and Therapeutics, Chinese University
of Hong Kong, Prince of Wales Hospital, Shatin, New Territories,
Hong Kong, SAR, People's Republic of China
e-mail: rcwma@cuhk.edu.hk

R. C. W. Ma · J. C. N. Chan
Hong Kong Institute of Diabetes and Obesity, Chinese University
of Hong Kong, Hong Kong, SAR, People's Republic of China

R. C. W. Ma · J. C. N. Chan
Li Ka Shing Institute of Life Sciences, Chinese University of Hong
Kong, Hong Kong, SAR, People's Republic of China

Methods We performed a meta-analysis of three GWAS comprising 684 patients with type 2 diabetes and 955 controls of Southern Han Chinese descent. We followed up the top signals in two independent Southern Han Chinese cohorts (totalling 10,383 cases and 6,974 controls), and performed in silico replication in multiple populations.

C. Hu · R. Zhang · C. Wang · J. Wang · W. Yu · F. Jiang · X. Ma ·
X. Hou · S. Tang · J. Lu · K. Xiang · Y. Bao · W. Jia (✉)
Department of Endocrinology and Metabolism, Shanghai Diabetes
Institute, Shanghai Key Laboratory of Diabetes Mellitus, Shanghai
Clinical Center for Diabetes, Shanghai Key Clinical Center for
Metabolic Disease, Shanghai Jiao Tong University Affiliated Sixth
People's Hospital, 600 Yishan Road, Shanghai 200233, People's
Republic of China
e-mail: wpjia@sjtu.edu.cn

C. Hu · X. Zhang
Shanghai Jiao Tong University Affiliated Sixth People's Hospital
South Campus, Shanghai, People's Republic of China

Springer

Results We identified *CDKN2A/B* and four novel type 2 diabetes association signals with $p<1\times10^{-5}$ from the meta-analysis. Thirteen variants within these four loci were followed up in two independent Chinese cohorts, and rs10229583 at 7q32 was found to be associated with type 2 diabetes in a combined analysis of 11,067 cases and 7,929 controls ($p_{\mathrm{meta}}=2.6\times10^{-8}$; OR [95% CI] 1.18 [1.11, 1.25]). In silico replication revealed consistent associations across multiethnic groups, including five East Asian populations ($p_{\mathrm{meta}}=2.3\times10^{-10}$) and a population of European descent ($p=8.6\times10^{-3}$). The rs10229583 risk variant was associated with elevated fasting plasma glucose, impaired beta cell function in controls, and an earlier age at diagnosis for the cases. The novel variant lies within an islet-selective cluster of open regulatory elements. There was significant heterogeneity of effect between Han Chinese and individuals of European descent, Malaysians and Indians.

Conclusions/interpretation Our study identifies rs10229583 near *PAX4* as a novel locus for type 2 diabetes in Chinese and other populations and provides new insights into the pathogenesis of type 2 diabetes.

Keywords Chinese · Diabetes · East Asians · Genetics · Genome-wide association study

T. F. Leung · H. Y. Sy · K. L. Hon
Department of Paediatrics, Chinese University of Hong Kong,
Hong Kong, People's Republic of China

G. N. Thomas
Department of Public Health, Epidemiology and Biostatistics,
University of Birmingham, Birmingham, UK

M. J. Go · Y. J. Kim · J. Y. Lee
Center for Genome Science, National Institute of Health,
Osong Health Technology Administration Complex,
Gangoe-myeon, Yeonje-ri, Cheongwon-gun, Chungcheongbuk-do,
Republic of Korea

K. Hara · H. Fujita · T. Yamauchi · T. Kadowaki
Department of Diabetes and Metabolic Diseases, Graduate School
of Medicine, University of Tokyo, Tokyo, Japan

K. Hara
Department of Integrated Molecular Science on Metabolic
Diseases, University of Tokyo Hospital, Tokyo, Japan

X. Sim
Centre for Molecular Epidemiology, Saw Swee Hock School
of Public Health, National University of Singapore, Singapore,
Republic of Singapore

X. Sim
Center for Statistical Genetics and Department of Biostatistics,
University of Michigan, Ann Arbor, MI, USA

Abbreviations

AAD	Age at diagnosis
ALR	Alternating logistic regressions
DIAGRAM Consortium	Diabetes Genetics Replication And Meta-analysis Consortium
eQTL	*cis*-Expression quantitative trait loci
FPG	Fasting plasma glucose
GWAS	Genome-wide association studies
HOMA-B	HOMA of beta cell function
HWE	Hardy–Weinberg equilibrium
LCL	Lymphoblastoid cell lines
LD	Linkage disequilibrium
MAF	Minor allele frequency
MC	Monte Carlo
The MuTHER Consortium	The Multiple Tissue Human Expression Resource Consortium
SNP	Single-nucleotide polymorphisms

H. Li · L. Lu · Y. Wang · X. Lin
Key Laboratory of Nutrition and Metabolism, Institute for
Nutritional Sciences, Shanghai Institutes for Biological Sciences,
Chinese Academy of Sciences and Graduate School of the Chinese
Academy of Sciences, Shanghai, People's Republic of China

J. W. Li · T. F. Chan
School of Life Sciences, Chinese University of Hong Kong,
Hong Kong, SAR, People's Republic of China

D. P. Ng
Saw Swee Hock School of Public Health, National University
of Singapore, Singapore, Republic of Singapore

K. Panoutsopoulou · A. G. Day-Williams · E. Zeggini
Wellcome Trust Sanger Institute,
Wellcome Trust Genome Campus,
Hinxton, Cambridge, UK

Y.-J. Fang
Department of Colorectal Surgery, State Key Laboratory
of Oncology in South China, Sun Yat-sen University
Cancer Center, Guangzhou, People's Republic of China

S. K. W. Tsui
School of Biomedical Sciences, Chinese University of Hong Kong,
Hong Kong, SAR, People's Republic of China

P. C. Leung
Department of Orthopaedics, Chinese University of Hong Kong,
Hong Kong, SAR, People's Republic of China

N. L. S. Tang
Department of Chemical Pathology, Chinese University of Hong
Kong, Hong Kong, SAR, People's Republic of China

Introduction

Type 2 diabetes is a common complex disease characterised by deficient insulin secretion and decreased insulin sensitivity. In 2010, 285 million people worldwide were affected by type 2 diabetes [1], with 60% of them located in Asia [2, 3]. China now has the largest number of patients with diabetes in the world, with an estimated 92 million affected individuals, and an additional 150 million with impaired glucose tolerance [4].

To identify common type 2 diabetes susceptibility variants, large-scale genome-wide association studies (GWAS) have been conducted in white individuals, yielding more than 60 genetic loci to date [5, 6]. Although many of these regions have been successfully replicated in Asian populations [7–11], discrepancies in allelic frequencies and effect sizes have demonstrated that interethnic differences exist. GWAS conducted in Japanese individuals [12, 13], as well as meta-analyses of GWAS in South Asian [14] and East Asian [15] groups, have revealed additional variants not detected in GWAS with white individuals, with several signals, including *KCNQ1*, later replicated in many populations [12, 13]. Previous GWAS in Chinese suggested several loci but lacked large-scale replication [16–18].

We therefore conducted this study to identify new type 2 diabetes susceptibility loci in Southern Han Chinese individuals. We performed a meta-analysis of three GWAS comprising 684 patients with type 2 diabetes and 955 controls, and analysed 2.9 million (genotyped and imputed) single-nucleotide polymorphisms (SNPs) in an additive model. Putatively associated SNPs ($p < 1 \times 10^{-5}$) were genotyped de novo in two independent Southern Han Chinese cohorts (10,383 cases and 6,974 controls), and SNPs reaching a genome-wide significance of $p < 5 \times 10^{-8}$ were replicated in silico in five East Asian and three non-East Asian populations for a total of 31,541 cases and 60,344 controls.

Methods

Participants In the first-stage discovery cohort (stage 1), we performed genome-wide scanning in three different case–control samples: 198 Hong Kong Chinese individuals (99 patients with type 2 diabetes and 99 healthy controls) in Hong Kong GWAS 1, 1,047 Hong Kong Chinese individuals (388 with type 2 diabetes and 659 controls) in Hong Kong GWAS 2 and 394 Shanghai Chinese (197 patients

J. Liu
Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore, Republic of Singapore

T. Y. Wong
Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Republic of Singapore

T. Y. Wong
Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Republic of Singapore

T. Y. Wong
Centre for Eye Research Australia, University of Melbourne, East Melbourne, VIC, Australia

S. Maeda
Laboratory for Endocrinology and Metabolism, RIKEN Center for Genomic Medicine, Yokohama, Japan

S. S. Cherny
Department of Psychiatry and State Key Laboratory of Brain and Cognitive Sciences, University of Hong Kong, Hong Kong, SAR, People's Republic of China

M. C. Y. Ng
Center for Genomics and Personalized Medicine Research, Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, NC, USA

A. P. Morris · S. Keildson · M. I. McCarthy
Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK

R. Hu
Institute of Endocrinology and Diabetology, Huashan Hospital, Shanghai Medical College, Fudan University, Shanghai, People's Republic of China

L. Ji
Department of Endocrinology and Metabolism, Peking University People's Hospital, Beijing, People's Republic of China

Y. S. Cho
Department of Biomedical Science, Hallym University, Chuncheon, Gangwon-do, Republic of Korea

E. S. Tai
Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Republic of Singapore

E. S. Tai
Graduate Medical School, Duke-National University of Singapore, Singapore, Republic of Singapore

M. I. McCarthy
Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Churchill Hospital, Oxford, UK

L. Baum
School of Pharmacy, Chinese University of Hong Kong, Hong Kong, SAR, People's Republic of China

with type 2 diabetes and 197 normal controls) in the Shanghai GWAS. Individuals included in the stage 2 replication included 5,366 with type 2 diabetes and 2,474 controls from Hong Kong, and 4,035 cases and 3,964 controls from Shanghai. We also included 325 cases and 368 controls from 178 Hong Kong families, as well as 657 cases and 168 controls from 248 Shanghai families.

Case–control samples for in silico replication in stage 3 were taken from several published type 2 diabetes GWAS in East Asian individuals. These included the Korea Association Resource Study [19], the Singapore Chinese from the Singapore Diabetes Cohort Study and the Singapore Prospective Study Program [20], the BioBank Japan Study [13] and a Han Chinese Study [21]. For stage 4 in silico replication in other populations, Malaysian participants from the Singapore Malay Eye Study, Indian participants from the Singapore Indian Eye Study [20] and participants of European descent in the Diabetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium [6] were included.

The study design, type 2 diabetes diagnostic criteria and clinical evaluation used in each study are described in the electronic supplementary material (ESM) Methods. The clinical characteristics of the study individuals are described in Table 1. Each study obtained approval from the appropriate institutional review boards of the respective institutions, and written informed consent was obtained from all participants. The overall study design is depicted in Fig. 1.

Quality control on the samples for the GWAS In our study, individuals were excluded from further analysis if: (1) duplicate samples existed; (2) the sex identified from the X chromosome was discordant with the sex obtained from the medical records; (3) the genotype call rate yield was <98%. We detected possible familial relationship using estimates of identity by descent derived from pair-wise analyses of independence ($r^2 \approx 0$) and quality SNPs. Individuals with evidence for relatedness were excluded ($\widehat{p_1} > 0.05$). ESM Table 1 shows the quality control for the participants in stage 1.

To discriminate individuals from different geographical origins, we conducted multidimensional scaling analysis using the genotype data obtained from unrelated individuals in the present study and the other 11 populations studied by the HapMap project (ESM Fig. 1). Individuals were excluded from subsequent analyses if they lay between clusters.

Genotyping and quality control on the SNP data Individuals for the stage 1, 3 and 4 analyses were genotyped using high-density SNP typing arrays that covered the entire genome. Only autosomal SNPs were included. Quality checks for SNPs were performed in the case and control samples separately, although the same criteria were applied to each. SNPs were excluded from further analysis if: (1) $p < 1 \times$ 10^{-4} for Hardy–Weinberg equilibrium (HWE); (2) minor allele frequency (MAF) was <1%; (3) call rate was <95%; in particular, SNPs with MAF≥1% but ≤5% were excluded if their call rate was <99%; or (4) the SNPs showed a significant difference in MAF ($p < 1 \times 10^{-4}$) between the Hong Kong control cohorts with other conditions (450 with epilepsy, 110 with eczema and 99 non-hypertensive individuals). Only SNPs that passed the quality control criteria for both cases and controls were used for further analysis. ESM Table 2 shows the quality control of the genotyping results in stage 1. We imputed genotypes for autosomal SNPs according to the 1000 Genomes reference panel. See the ESM Methods for further details.

For de novo replication in stage 2, all selected SNPs were genotyped in the Hong Kong and Shanghai case–control samples by a primer extension of multiplex products with detection by Matrix-assisted laser desorption ionisation-time of flight mass spectroscopy using a MassARRAY platform (Sequenom; San Diego, CA, USA). Family samples were genotyped using TaqMan SNP Genotyping Assays (Applied Biosystems, Foster City, CA, USA) or by direct sequencing.

Statistical analysis All statistical analyses were performed using PLINK version 1.07 (http://pngu.mgh.harvard.edu/~purcell/plink/) [22], SAS version 9.1 (SAS Institute, Cary, NC, USA) or SPSS for Windows version 18 (SPSS, Chicago, IL, USA), unless specified otherwise. Haploview version 4.1 was used to generate pair-wise linkage disequilibrium (LD) measures (r^2).

To test for an association with type 2 diabetes, we applied logistic regression under an additive genetic model using the MACH2DAT software (www.sph.umich.edu/csg/abecasis/MACH/download/) [23] adjusted for sex and age according to situations in the individual studies.

To combine the type 2 diabetes association results in stage 1, GWAMA software (www.well.ox.ac.uk/gwama/) [24] was used to calculate the combined estimates of the ORs (95% CIs) from multiple groups by weighting the natural log-transformed ORs of each study using the inverse of their variance under the random effect model [25]. By using the random effect model, we excluded SNPs with some degree of heterogeneity between studies, which helped to attenuate the number of false-positive findings in this study. Cochran's Q statistic ($p < 0.05$) and I^2 index were used to assess the heterogeneity of ORs between studies.

The most strongly associated SNPs were prioritised for follow-up in stage 2 based on the meta-analysis results from stage 1. SNPs located within a previously reported type 2 diabetes locus were excluded. We finally considered 13 top and proxy SNPs from four distinct loci available in all three GWAS with (1) a meta-analysis $p < 1 \times 10^{-5}$; (2) a heterogeneity test $p > 0.05$; (3) the same direction of risk allele across all three GWAS; (4) a common allele frequency (MAF≥0.1).

Table 1 Clinical characteristics of the participants

Study	Cohort	N (male %)	Age (years)	AAD (year)	Diabetes duration (years)	BMI (kg/m²)	FPG (mmol/l)
Stage 1 (genome scan)							
HK1	Control	99 (36.4)	37.3±10.2	–	–	20.8±2.0	4.7±0.4
	T2D patient	99 (40.4)	40.6±8.8	31.8±7.7	8.0±8.3	30.9±4.4	–
HK2	Diseased control	659 (48.7)	37.1±17.0	–		23.3±3.7	–
	T2D patient	388 (49.5)	60.6±10.8	51.1±12.1	9.5±7.0	25.0±3.8	–
SH	Control	197 (50.8)	66.4±10.1	–	–	20.6±1.7	4.8±0.4
	T2D patient	197 (57.9)	41.6±10.4	34.5±4.8	7.3±8.5	23.8±4.1	–
Stage 2 (de novo replication in Chinese)							
HK1	Adolescent control	985 (44.2)	15.5±1.9	–	–	22.7±5.4	4.9±0.4
	Adults control	513 (47.0)	42.0±10.4	–	–	19.9±3.5	4.7±0.3
	Elderly control	976 (51.4)	72.3±5.3	–	–	23.2±3.2	–
	T2D patient	5,366 (45.1)	56.7±13.4	48.8±14.9	6.6±6.9	24.6±5.3	–
SH1	Control	3,964 (37.6)	51.3±13.5	–	–	23.6±3.2	5.0±0.5
	T2D patient	4,035 (52.0)	61.2±12.1	54.2±11.3	7.2±6.9	24.5±3.5	–
HK Family 2	Control	368 (41.0)	37.0±13.6	–	–	24.0±4.1	4.9±0.4
	T2D patient	325 (40.6)	48.0±14.4	41.7±13.1	6.3±7.6	25.9±4.4	–
SH Family 2	Control	168 (51.2)	62.8±11.2	–	–	23.7±3.5	4.8±0.6
	T2D patient	657 (43.7)	54.6±15.6	50.0±14.2	4.9±7.3	23.9±3.5	–
Stage 3 (in silico replication in East Asians)							
Japanese	Control	3,023 (54.5)	51.9±15.2	–	–	22.4±3.7	–
	T2D patient	4,465 (68.0)	65.8±10.0	56.5±11.4	9.4±8.4	24.1±3.8	–
Korean 1	Control	2,943 (46.0)	51.1±8.6	–	–	24.1±3.0	4.5+0.4
	T2D patient	1,042 (51.7)	56.4±8.6	–	–	25.5±3.3	7.0±2.6
Korean 2	Control	1,305 (54.5)	65.2±2.6	–	–	23.9±3.0	5.0±0.5
	T2D patient	1,183 (46.5)	58.6±7.1	–	–	25.2±3.4	7.4±2.7
Singapore Chinese 1	Control	1,006 (21.6)	47.7±11.1	–	–	22.3±3.7	4.7±0.4
	T2D patient	1,082 (37.2)	65.1±9.7	55.7±12.0	–	25.3±3.9	–
Singapore Chinese 2	Control	939 (63.8)	46.7±10.2	–	–	22.8±3.4	4.7±0.5
	T2D patient	928 (64.9)	63.7±10.8	52.2±14.4	–	25.4±3.8	–
Chinese	Control	1,839 (43.7)	54.1±9.2	–	–	24.00±3.18	5.04±0.35
	T2D patient	1,873 (46.0)	58.6±8.4	–	–	25.00±3.24	8.43±2.90
Stage 4 (in silico replication in non-East Asians)							
Singapore Malaysian	Control	1,240 (52.0)	56.9±11.4	–	–	25.1±4.8	–
	T2D patient	794 (51.0)	62.3±9.90	54.4±11.2	–	27.8±4.9	–
Singapore Indians	Control	1,169 (48.4)	55.7±9.7	–	–	25.3±4.4	–
	T2D patient	977 (54.4)	60.7±9.9	51.4±10.6	–	27.1±5.1	–
DIAGRAM+	Control	38,987 (–)	–	–	–	–	–
	T2D patient	8,130 (–)	–	–	–	–	–

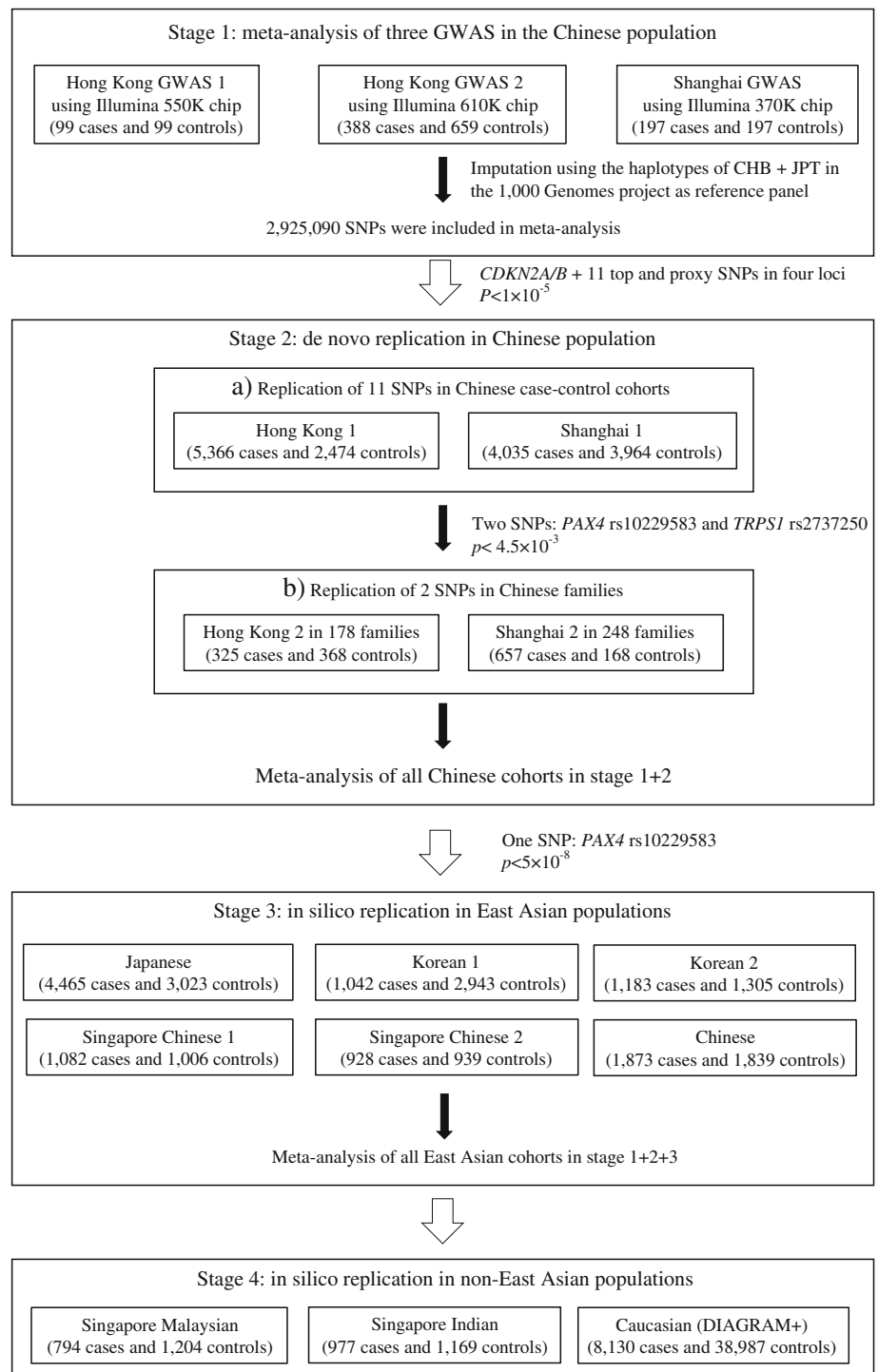Data are shown as N, percentage or mean ± SD

T2D, type 2 diabetes

For SNPs imputed across all three studies, we selected the most significant SNP associated with type 2 diabetes. ESM Tables 3 and 4 describe the details of the selected SNPs and the quality control for the genotyping results in stage 2.

In the replication stage, genotype frequencies were compared between cases and controls using logistic regression under an additive genetic model. In the family studies, alternating logistic regressions (ALR) implemented in the SAS procedure GENMOD was used to test for the association between type 2 diabetes and SNPs under an additive genetic model adjusted for age and sex. ALR is one type of generalised estimating equation applicable to binary outcomes that can handle correlated data (e.g. familial correlation). ORs (95% CIs) are presented in both analyses. Meta-

Fig. 1 Summary of study design. CHB, Han Chinese in Beijing, China; JPT, Japanese in Tokyo, Japan

Stage 1: meta-analysis of three GWAS in the Chinese population

| Hong Kong GWAS 1 using Illumina 550K chip (99 cases and 99 controls) | Hong Kong GWAS 2 using Illumina 610K chip (388 cases and 659 controls) | Shanghai GWAS using Illumina 370K chip (197 cases and 197 controls) |

Imputation using the haplotypes of CHB + JPT in the 1,000 Genomes project as reference panel

2,925,090 SNPs were included in meta-analysis

CDKN2A/B + 11 top and proxy SNPs in four loci $P < 1 \times 10^{-5}$

Stage 2: de novo replication in Chinese population

a) Replication of 11 SNPs in Chinese case-control cohorts

| Hong Kong 1 (5,366 cases and 2,474 controls) | Shanghai 1 (4,035 cases and 3,964 controls) |

Two SNPs: *PAX4* rs10229583 and *TRPS1* rs2737250 $p < 4.5 \times 10^{-3}$

b) Replication of 2 SNPs in Chinese families

| Hong Kong 2 in 178 families (325 cases and 368 controls) | Shanghai 2 in 248 families (657 cases and 168 controls) |

Meta-analysis of all Chinese cohorts in stage 1+2

One SNP: *PAX4* rs10229583 $p < 5 \times 10^{-8}$

Stage 3: in silico replication in East Asian populations

| Japanese (4,465 cases and 3,023 controls) | Korean 1 (1,042 cases and 2,943 controls) | Korean 2 (1,183 cases and 1,305 controls) |
| Singapore Chinese 1 (1,082 cases and 1,006 controls) | Singapore Chinese 2 (928 cases and 939 controls) | Chinese (1,873 cases and 1,839 controls) |

Meta-analysis of all East Asian cohorts in stage 1+2+3

Stage 4: in silico replication in non-East Asian populations

| Singapore Malaysian (794 cases and 1,204 controls) | Singapore Indian (977 cases and 1,169 controls) | Caucasian (DIAGRAM+) (8,130 cases and 38,987 controls) |

analyses and heterogeneity tests were conducted as described previously to combine estimates of the ORs (95% CIs) from multiple case–control and family groups under the fixed effect model. Multiple testing in the combined analysis of the case–control study were controlled by Bonferroni correction, and *p* $<4.5 \times 10^{-3}$ (0.05 divided by 11 SNPs in the stage 2 replication studies) was used as the threshold for filtering SNPs genotyped in the family studies.

Continuous data are presented as mean ± SD or geometric mean (95% CI). Traits were $\log_e$-transformed due to skewed distributions. Associations between genotypes and quantitative traits were tested by linear regression (adjusted for sex, age and/or BMI) in each healthy control cohort, as were associations for age at diagnosis (AAD) among patients with type 2 diabetes (adjusted for sex, BMI and/or HbA$_{1c}$). Meta-analyses implemented by GWAMA were

applied to combine effect size ($\beta \pm$SE) from multiple groups under the fixed effect model.

We performed bioinformatics and *cis*-expression quantitative trait loci (eQTL) analysis for functional implication of the identified SNP. See the ESM Methods for additional information on methods, including adjustment for genomic control and the gene network analysis.

Results

Meta-analysis of patients with Chinese ancestry A summary of the study design and the clinical characteristics of the participants in all stages are shown in Fig. 1 and Table 1. In stage 1, we genotyped 684 patients with type 2 diabetes and 955 controls. We did not detect any population stratification between case and control individuals in multidimensional scaling analysis for all GWAS (ESM Fig. 2). Meta-analysis was implemented to combine the individual association results for 2,925,090 imputed and genotyped SNPs (under additive genetic models) available in all three GWAS using the inverse-variance approach for random effect models.

In the stage 1 meta-analysis of three Chinese GWAS, 44 SNPs within five loci were prioritised for follow-up (Fig. 2 and ESM Table 5). We did not observe a substantial change in the stage 1 results after adjusting either for λs (1.01–1.04 in individual cohorts) or the first principal component in the meta-analysis, reflecting that the results were not likely to be due to population stratification (ESM Fig. 3 and ESM Table 6).

Of the five loci identified in stage 1, *CDKN2A/B* has previously been reported to be strongly associated with type 2 diabetes. In line with our previous findings, two SNPs in *CDKN2A/B* showing strong signals for type 2 diabetes in the present study were in high LD ($r^2 \approx 0.8$) with rs10811661, which is well-replicated in most populations. After eliminating the signal of *CDKN2A/B* and redundant markers, we took forward 13 top and proxy SNPs among the remaining 42 SNPs in four regions to stage 2, de novo replication, in two independent Chinese case–control cohorts (ESM Table 3). We

successfully obtained genotypes for 11 SNPs in Hong Kong replication 1 cohort with 5,366 cases and 2,474 controls, and Shanghai replication 1 cohort with 4,035 cases and 3,964 controls to proceed for subsequent analysis (ESM Table 4). Of these, rs10229583 and rs2737250, located on chromosomes 7 and 8, respectively, gave $p \leq 4.5 \times 10^{-3}$ (threshold of significance after Bonferroni correction) with the same directions of association as the original signals (Table 2). These two SNPs were genotyped in 1,518 additional samples from 426 families of Han Chinese descent (325 cases and 368 controls from 178 Hong Kong families, and 657 cases and 168 controls from 248 Shanghai families). Although we did not detect a significant association in either family study using ALR, all were in the concordant direction for rs10229583 (ESM Table 7). Taken together, the overall observed association for type 2 diabetes with rs10229583 by combining all studies from Chinese ancestry in stages 1 and 2 yielded an OR (95% CI) of 1.18 (1.11, 1.25) with a corresponding $p=2.6 \times 10^{-8}$ (Table 3). For another variant taken to genotyping in family samples, rs2737250, meta-analysis of GWAS and de novo genotyping in the Hong Kong and Shanghai case–control samples revealed OR 1.10 (1.05, 1.15) with a corresponding $p=7.05 \times 10^{-5}$ using a fixed effect model (p for heterogeneity test=0.0012, I^2=0.852), with OR 1.16 (1.01, 1.33), p=0.0299 by random effect model. However, genotyping of the variant in the Hong Kong and Shanghai family samples suggested an association in the opposite direction (ESM Table 7).

Meta-analysis in East Asian and other populations To further validate the association of rs10229583 with type 2 diabetes, we conducted in silico replication of rs10229583 in five East Asian GWAS (one Japanese, two Korean, one Singapore Chinese and one Han Chinese study), and three non-East Asian GWAS (Singapore Indian, Singapore Malaysian and the DIAGRAM Consortium). Meta-analysis for the East Asian populations ($p=2.3 \times 10^{-10}$) gave an OR (95% CI) of 1.14 (1.09, 1.19). Among non-East Asian populations, we observed replication of the association in participants of European descent from the DIAGRAM Consortium ($p=8.6 \times 10^{-3}$),
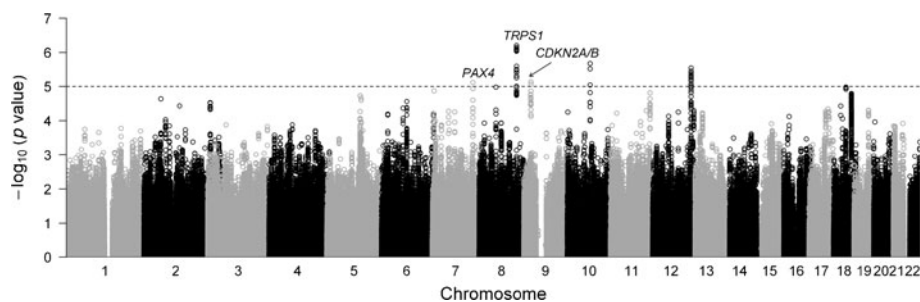

Fig. 2 Manhattan plot of combined genome-wide association results from the Hong Kong 1, Hong Kong 2 and Shanghai studies based on the random effect models. The *y*-axis represents the $-\log_{10} p$ value, and the *x*-axis represents the 2,925,090 analysed SNPs. The dashed horizontal line indicates the threshold of significance $p<1 \times 10^{-5}$. There are 44 points with $p<1 \times 10^{-5}$, and the arrow and labels localise the susceptibility loci to type 2 diabetes uncovered in the present study

Table 2 Association results for type 2 diabetes (T2D) with 11 top and proxy SNPs in de novo replication stage in Chinese populations

SNP	Chromosome	Nearest gene(s)	Position (B36)	Minor/ major allele	Hong Kong replication 1 (5,366 T2D vs 2,474 controls)				Shanghai replication 1 (4,035 T2D vs 3,964 controls)				Combined			
					Case MAF	Control MAF	OR (95% CI)	$p_{additive}$	Case MAF	Control MAF	OR (95% CI)	$p_{additive}$	OR (95% CI)	p_{meta} (uncorrected)	p_{het}	I^2
rs10229583	7	PAX4	127034139	A/G	0.847	0.83	1.14 (1.03, 1.23)	0.0077	0.846	0.825	1.16 (1.08, 1.27)	3.7×10^{-4}	1.15 (1.08, 1.22)	1.0×10^{-5}	0.6406	0.000
rs2721960	8	TRPS1	116725904	T/C	0.657	0.644	1.05 (0.98, 1.14)	0.1566	0.655	0.638	1.08 (1.01, 1.15)	0.0277	1.06 (1.02, 1.12)	0.0095	0.7067	0.000
rs2737250	8	TRPS1	116731048	G/A	0.631	0.62	1.05 (0.98, 1.12)	0.1807	0.641	0.621	1.09 (1.02, 1.16)	0.0090	1.08 (1.02, 1.12)	0.0045	0.4582	0.000
rs3858158	10	COL13A1	71310056	C/T	0.516	0.521	0.98 (0.92, 1.05)	0.6000	0.569	0.561	1.03 (0.97, 1.10)	0.3211	1.01 (0.96, 1.05)	0.7408	0.3026	0.506
rs2395272	10	COL13A1	71310261	A/G	0.531	0.534	0.99 (0.93, 1.06)	0.7680	0.594	0.584	1.04 (0.97, 1.11)	0.2312	1.02 (0.97, 1.06)	0.4589	0.3027	0.364
rs57703465	10	COL13A1	71311074	T/C	0.654	0.662	0.96 (0.89, 1.04)	0.3463	0.667	0.656	1.05 (0.98, 1.12)	0.1502	1.01 (0.96, 1.06)	0.6467	0.0976	0.765
rs11065441	12	P2RX7	120045354	C/T	0.728	0.724	1.02 (0.94, 1.11)	0.6224	0.728	0.733	0.97 (0.91, 1.04)	0.4312	0.99 (0.94, 1.05)	0.7756	0.3748	0.000
rs684201	12	P2RX7	120054726	A/G	0.73	0.726	1.02 (0.94, 1.10)	0.5916	0.735	0.739	0.98 (0.91, 1.05)	0.5609	1.00 (0.94, 1.05)	0.9462	0.4332	0.000
rs11065450	12	P2RX7	120064040	A/C	0.682	0.688	0.97 (0.90, 1.05)	0.4995	0.702	0.707	0.98 (0.92, 1.05)	0.5520	0.98 (0.93, 1.03)	0.3699	0.9111	0.000
rs208290	12	P2RX7	120078439	T/C	0.612	0.609	1.01 (0.94, 1.09)	0.7086	0.643	0.645	0.99 (0.93, 1.06)	0.7950	1.00 (0.95, 1.05)	0.9605	0.6472	0.000
rs10849851	12	P2RX7	120081027	G/A	0.727	0.72	1.03 (0.95, 1.12)	0.4079	0.737	0.741	0.98 (0.91, 1.05)	0.5237	1.00 (0.95, 1.05)	0.9308	0.3002	0.068

Nearest Entrez genes within 250 kb

p, p_{meta} and p_{het} represent p values from logistic regression without any adjustment under the additive genetic model, meta-analysis under a fixed effect model (uncorrected for multiple testing) and test of heterogeneity, respectively

ORs are reported with respect to the minor allele

Table 3 Association results for rs10229583 and type 2 diabetes (T2D)

Stage	Cohort	Adjustment	N		Risk allele frequencies		OR (95% CI)	$p_{additive}$ (uncorrected GC)	p_{het}	I^2
			T2D	Control	T2D	Control				
1. Discovery	Hong Kong GWAS 1	Sex and age	99	99	0.879	0.818	1.48 (0.85, 2.59)	0.1645		
	Hong Kong GWAS 2	Sex and age	388	659	0.857	0.820	1.56 (1.14, 2.13)	0.0055		
	Shanghai GWAS	None	197	197	0.873	0.777	1.92 (1.32, 2.79)	5.0×10^{-4}		
	Meta-analysis of GWAS		684	955			1.66 (1.33, 2.07)	7.7×10^{-6}	0.6455	0.000
2. De novo replications in Hong Kong and Shanghai	Hong Kong replication 1	None	5,366	2,474	0.847	0.831	1.13 (1.03, 1.24)	7.7×10^{-3}		
	Shanghai replication 1	None	4,035	3,964	0.846	0.825	1.17 (1.07, 1.27)	3.7×10^{-4}		
	Hong Kong family replication 2	Sex and age	325	368	0.872	0.856	1.22 (0.85, 1.74)	0.2817		
	Shanghai family replication 2	Sex and age	657	168	0.824	0.813	1.09 (0.80, 1.49)	0.5757		
	Replication in Chinese		10,383	6,974			1.15 (1.08, 1.22)	1.0×10^{-5}	0.6406	0.000
	Meta-analysis of Chinese		11,067	7,929			1.18 (1.11, 1.25)	2.6×10^{-8}	0.0839	0.596
3. In silico replications in East Asians	Japanese replication	None	4,465	3,023	0.892	0.881	1.11 (1.01, 1.23)	0.0379		
	Korean replication 1	None	1,042	2,943	0.894	0.878	1.17 (0.99, 1.38)	0.0577		
	Korean replication 2	None	1,183	1,305	0.841	0.844	0.98 (0.84, 1.15)	0.8101		
	Singapore Chinese replication 1	None	1,082	1,006	0.832	0.819	1.07 (0.95, 1.20)	0.2728		
	Singapore Chinese replication 2	None	928	939	0.833	0.816				
	Chinese replication	First 2 PCs	1,873	1,839	0.8396	0.8167	1.17 (1.04, 1.32)	0.01091		
	Replication in other East Asian		10,573	11,055			1.10 (1.04, 1.17)	6.0×10^{-4}	0.6767	0.000
	Meta-analysis of East Asian		21,640	18,984			1.14 (1.09, 1.19)	2.3×10^{-10}	0.5939	0.000
4. In silico replications in South Asians and Europeans	Singapore Malaysian replication	None	794	1,204	0.798	0.804	0.97 (0.83, 1.14)	0.7185		
	Singapore Indian replication	None	977	1,169	0.647	0.682	0.86 (0.76, 0.98)	0.0276		
	DIAGRAM	None	8,130	38,987	–	–	1.06 (1.02, 1.12)	8.6×10^{-3}		
	Replication in non-East Asian		9,901	41,360			1.03 (0.99, 1.08)	0.1156	0.0042	0.878

ORs and 95% CIs were reported with respect to the T2D-related risk alleles (G)

p_{het} refers to the p value obtained from the heterogeneity test

GC, genomic control; PC, principal components
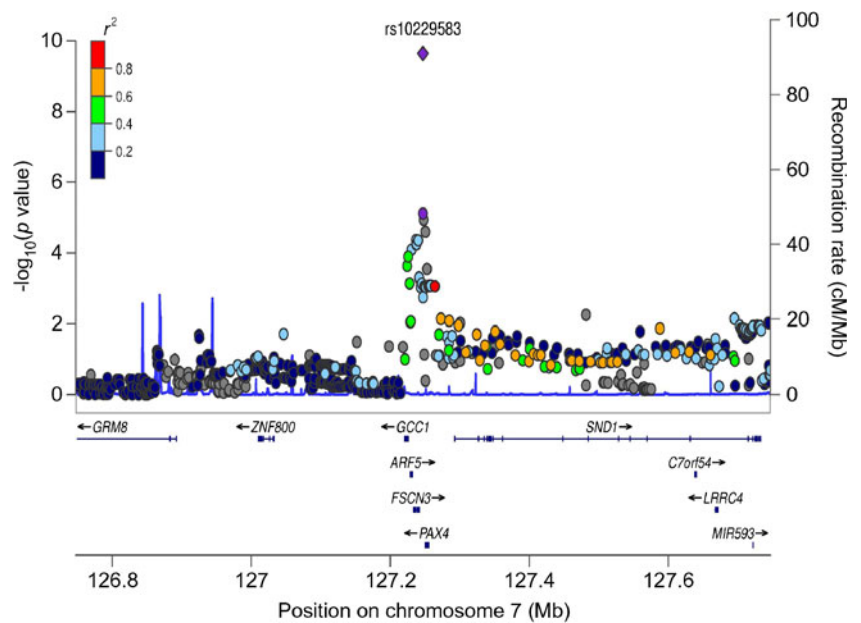
with OR 1.06 and (95% CI 1.02, 1.12) (Table 3 and Figs 3 and 4).

Impact of rs10229583 on clinical traits and course of disease We next investigated the associations of rs10229583 with the AAD of type 2 diabetes and quantitative metabolic traits related to type 2 diabetes. Among all the patients with type 2 diabetes, individuals who carried the common, type 2 diabetes risk allele (G) were concordantly and significantly younger at the time of diagnosis in both Hong Kong and Shanghai, and the meta-analysis showed that presence of the risk variant had a significant association with younger AAD

Fig. 4 Forest plot for meta-analysis of the association between type 2 diabetes and rs10229583 for all populations in the present study. ORs and 95% CIs were reported with respect to the type 2 diabetes-related risk alleles (G)
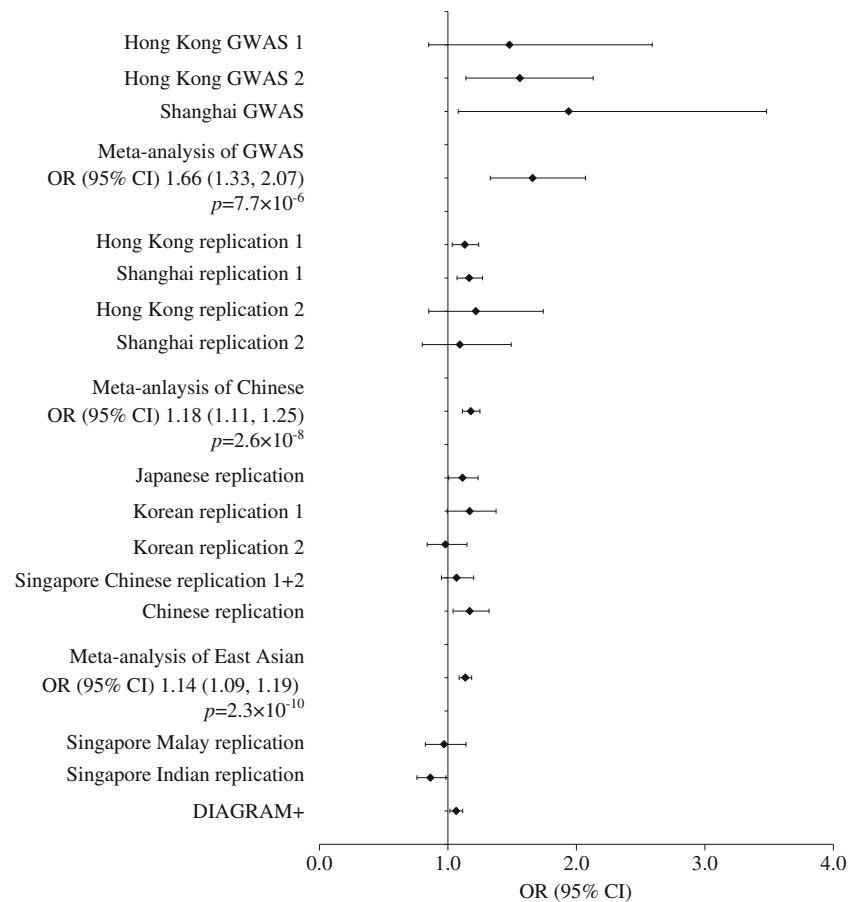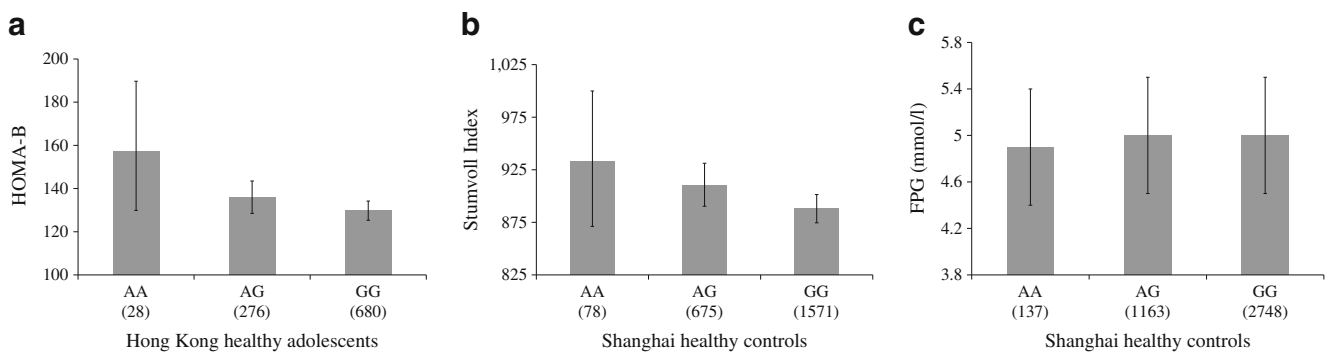
Fig. 5 Associations of the risk variant (G allele) of rs10229583 with measures of insulin secretion in Chinese controls. (**a**) Association with reduced HOMA-B in a Hong Kong Chinese adolescent cohort ($p=0.0221$). (**b**) Association with a reduced Stumvoll Index of beta cell function in healthy Shanghai controls ($p=0.0303$). (**c**) Association of the risk variant with higher FPG in healthy Shanghai controls ($p=0.0460$). Data are expressed as mean (for FPG) or geometric mean (for HOMA-B and Stumvoll Index). SDs or 95% CIs are expressed as error bars. The number of individuals analysed for each genotype is shown in parentheses under each column

($p=2.3\times10^{-4}$, $\beta_{\text{unadjusted}}\pm\text{SE}=-0.90\pm0.24$), which remained unchanged following adjustment for sex and BMI (ESM Table 8). We also observed a nominal association of the G-alleles of rs10229583 with beta cell function as assessed by HOMA-B ($\beta_{\text{unadjusted}}\pm\text{SE}=-0.06\pm0.03$, $p=0.0221$) in healthy Hong Kong adolescents, a reduced Stumvoll Index ($\beta_{\text{unadjusted}}\pm\text{SE}=0.03\pm0.01$, $p=0.0303$) and increased fasting plasma glucose (FPG) level ($\beta_{\text{unadjusted}}\pm\text{SE}=0.03\pm0.01$, $p=0.0460$) in healthy Shanghai adults (Fig. 5).

Functional implication of the identified locus rs10229583 In order to evaluate the functional implication of our identified variant, we performed an extensive bioinformatics analysis. Consistent with its observed effect on pancreatic beta cell function, the gene region of our locus has been identified as one of the islet-selective clusters of open regulatory elements using a formaldehyde-assisted isolation of regulatory elements coupled with high-throughput sequencing in human pancreatic islets [26]. In addition, the variant and its tagging SNPs lie within an area near *PAX4* and *SND1*, which is enriched with DNase I hypersensitive sites, histone H3 lysine modifications and CCCTC factor binding in human islets (ESM Fig. 4) [27].

We next investigated the relationship of rs10229583 with eQTLs in adipose tissue and other tissues in available datasets. The variant rs1440971, a proxy of our associated SNP (MAF~0.1, $r^2=0.8$ and D'=1, to rs10229583), was significantly associated with the level of expression of *GRM8*, *ARF5* and *PAX4* in lymphoblastoid cells in the GenCord Project, although this did not correlate with the eQTL peak (ESM Fig. 5 and ESM Table 9) [28]. Analysis of all eQTLs associated with rs10229583, or its close proxy, rs1440971, was performed using data from the Multiple Tissue Human Expression Resource (MuTHER) Consortium [29]. Of note, eQTL data were only available for *PAX4* in adipose tissue, but not lymphoblastoid cell lines (LCLs) or skin, for which no expression data were available from MuTHER. There was a

nominal association ($p<0.05$) between the variant and expression of *C7orf54* and *ARF5* in LCLs, and *C7orf68* in adipose tissue (ESM Table 10). The r^2 between the GWAS SNP and the peak eQTL SNPs ranged between 0.56 and 1.

Complex diseases such as type 2 diabetes are caused by a combination of alterations, and each genomic perturbation or alteration can potentially impact on thousands of genes [30, 31]. Nevertheless, functionally important genes often organise into the same pathway of functional grouping. Therefore, we can overlay the alterations on a gene network that was built using highly confident gene–gene relationships (ESM Fig. 6) [32]. We identified interactions between these genes, with additional interaction with other key pancreatic transcription factors such as *NEUGRO3* (ESM Fig. 6). Taken together, we speculate that *ARF5*, *GCC1*, *SND1* and *PAX4* may function together with *NEUGRO3* in the same network for pancreatic islet development.

Heterogeneity of effect in Chinese vs other ethnic groups To investigate why the novel loci identified in the present study had not been detected in previous GWAS performed in other populations, we examined the heterogeneity of effect between East Asians and Europeans. There was no evidence of heterogeneity of effect between Chinese, Korean and Japanese populations, but significant heterogeneity of effect was seen between Han Chinese and individuals of European descent in the DIAGRAM Consortium, as well as between Chinese, Malaysians and Indians (ESM Table 11).

To test for the variation of LD structure between Chinese and other populations, we implemented the targeted varLD approach to examine the pattern of r^2 between every pair of SNPs within the 100 kb region centred on our index SNP rs10229583. (www.statgen.nus.edu.sg/~SGVP/software/varld.html) [33]. This region shows highly significant evidence of LD variation between Chinese, European (Monte Carlo [MC] $p=0.0018$), and African (MC $p=0.0003$) individuals, but nominal evidence of variations between

Chinese and Japanese (MC $p=0.0107$) (ESM Table 12 and ESM Figs 7 and 8). We also observed discrepancies in allele frequency of rs10229583 between East Asians, Europeans and Africans (ESM Table 12).

Discussion

This study reports a meta-analysis of GWAS for type 2 diabetes in a Chinese population, and has identified a novel diabetes-associated locus. Furthermore, we replicated the association in additional East Asian samples, and found an association in samples of European descent. In addition to the multiethnic samples used in our study, our study also benefits from a detailed phenotyping of the Chinese samples, which allowed additional analyses of the effect of the risk variant on clinical traits and the course of disease to be carried out.

Type 2 diabetes in Asians is characterised by an earlier AAD, strong family history and evidence of impaired beta cell function [2, 3]. In a recent nationwide study conducted in China, the prevalence of diabetes was 3.2% among persons aged 20–39 years, and 11.5% among adults aged 40–59 [4]. The risk variant we identified, rs10229583, was associated with earlier AAD in both the Hong Kong and Shanghai samples, highlighting its potential contribution to young-onset diabetes in the Chinese population. Healthy adults and adolescents who carry the risk variant were found to have elevated fasting glucose and impaired beta cell function, respectively.

The novel locus for type 2 diabetes we identified, rs10229583, is located downstream of the *ARF5* and *PAX4* genes in 7q32, and upstream of *SND1*. *PAX4*, which is a member of the paired box family of transcription factors, plays a critical role in pancreatic beta cell formation during fetal development [34, 35] and is therefore a very strong candidate for the implicated gene. The gene region lies within an area of islet-specific cluster of open chromatin sites and may therefore act in *cis* with local chromatin and regulatory changes [25]. *PAX4* is expressed in early pancreatic endocrine cells, but expression is later restricted to beta cells and it is not expressed in mature pancreas [36]. In pancreatic endocrine cells, *PAX4* represses ghrelin and glucagon expression, and can induce the expression of *PDX1*, a key transcription factor for islet development [37]. Targeted disruption of *PAX4* in mice was found to lead to reduced beta cell mass at birth [37].

Several human studies have implicated *PAX4* in the pathogenesis of diabetes [38, 39]. In one report, a missense mutation (R121W) was identified in six heterozygous patients and one homozygous patient out of 200 unrelated Japanese patients with type 2 diabetes [39]. For example, Japanese patients carrying *PAX4* mutations have severe defects in first-phase insulin secretion [40]. Mutations in *PAX4*

may lead to rare monogenic forms of young-onset diabetes [41]. Common variants in several other MODY genes, namely, *HNF4a*, *HNF1a* and *TCF2*, have been identified as susceptibility loci for type 2 diabetes [42].

Our finding is consistent with other studies that have highlighted the important role of genes implicated in pancreatic development in the pathogenesis of type 2 diabetes. In a previous study, a risk variant at *HNF4a* has been found to be associated with increased risk of type 2 diabetes, and carriers of the risk allele have impaired beta cell function [43]. The MAF of the R121W *PAX4* mutation was 1% in Asians, and the mutation is in low LD with rs10229583. It is possible that both rare mutations and common variation within the same gene confer risk towards type 2 diabetes independently. The common variant we identified, rs10229583, may be associated with altered gene expression, while the other rare non-synonymous mutations lead to impaired gene function. For example, while common non-coding variants in *MTNR1B* increase type 2 diabetes risk with a modest effect, large-scale resequencing has identified rare loss-of-function *MTNR1B* variants that significantly contribute towards type 2 diabetes risk [44]. Some regulatory elements harbouring type 2 diabetes-associated loci have recently been found to exhibit allele-specific differences in activity, providing evidence supporting the functional role of non-coding common variants identified through GWAS [26, 27].

The recent East Asian meta-analysis comprising eight type 2 diabetes GWAS identified a locus on chromosome 7 near *GRIP* and *GCC1-PAX4* to be associated with type 2 diabetes. The protein encoded by *GCC1* may play a role in transmembrane transport [45]. The variant identified from the East Asian study, rs6467136, appears to be independent of our signal, with $r^2=0.044$ in our Chinese samples (ESM Fig. 8). Furthermore, we found no change in the effect size of rs10229583 after conditioning on rs6467136 (OR [95% CI]=1.20 [1.11, 1.29], $p=4.6\times10^{-6}$ vs OR [95% CI]=1.19 [1.10, 1.29], $p=1.6\times10^{-5}$, before and after the conditional analysis in 9,886 Chinese samples). Likewise, rs6467136 had little change in effect after conditioning on rs10229583 (OR [95% CI]=1.09 [1.02, 1.17], $p=0.0125$ before; OR [95% CI]=1.07 [0.99, 1.15], $p=0.0729$ after). In the recent analysis from the DIAGRAM Consortium, rs231362 near *KCNQ1* was identified to be associated with type 2 diabetes. This signal is independent of the original signal identified in the Japanese population as revealed by conditional analysis. Consistent with the evidence observed for *KCNQ1*, our finding highlighted that multiple common genetic variations within the same gene region may independently contribute to disease risk [6, 12]. It will be worthwhile undertaking a further investigation of this region to search for population-specific and/or disease causal variants in different ethnic groups by fine-mapping as well as transethnic mapping.

The other genes in the region of our identified variant are also potential candidate genes for diabetes. *ARF5* belongs to a family of guanine nucleotide-binding proteins that have been shown to play a role in vesicular trafficking and as activators of phospholipase D [46]. Islet expression of *ARF5* was found to be induced threefold in rats receiving a high-carbohydrate diet [47]. The nearby *SND1* gene, also known as the p100 transcription co-activator, is a member of the micronuclease family and plays a key role in transcription and splicing. The p100 transcriptional co-activator is present in endocrine cells and tissues, including the pancreas of cattle [48].

Among the type 2 diabetes loci first identified in non-European populations, other than *KCNQ1*, few have consistently been found to show a significant association in studies of individuals of European descent [6, 14, 15, 42]. The diabetes gene variant we identified, rs10229583, also showed a significant association in Europeans in the DIA-GRAM Consortium, with a smaller effect size compared with East Asian individuals ($p=0.0024$ by Cochran's Q statistics, $I^2=0.8913$). Interestingly, rare *PAX4* mutations were first identified in Asian MODY probands [39, 41], but seldom found in those of European descent [49, 50]. This suggests that *PAX4*, like *KCNQ1*, may be particularly relevant for the pathogenesis of type 2 diabetes in East Asians individuals. Interestingly, rs10229583 is also in strong LD with a region spanning the neighbouring *SND1* gene (Fig. 3). Further resequencing and transethnic mapping should help to identify the causal gene variant for type 2 diabetes within this region.

The novel locus we identified in Chinese individuals with type 2 diabetes has not been detected in previous GWAS performed in mainly individuals of European descent. We noted a highly significant LD variation between Chinese and European individuals in the region surrounding our identified variant. There is also significant variation in allele frequencies in Chinese compared with Europeans, as well as between Chinese and African individuals. This ethnic difference in LD pattern and risk allele frequency may lead to a differential impact in different populations and warrants further investigation by resequencing.

Our study has several limitations. The sample size of our GWAS was modest, resulting in limited power to identify genetic variants with small effect sizes. We have limited our discovery study to Southern Han Chinese, although the consistent replication seen in other East Asian population suggests that the findings may be applicable to other populations of Chinese descent.

In summary, we identify rs10229583 near *PAX4* as a novel locus for type 2 diabetes in Chinese and other populations, providing new insights into the pathogenesis of type 2 diabetes.

Duality of interest The authors declare that there is no duality of interest associated with this manuscript.

Contribution statement The study was supervised by RCWM, CH, WYS, YB, JCNC and WJ. The experiments were conceived and designed by RCWM, CH, PK, TFL, GNT, MJG, KH, TW, J-YL, Y-JF, SM, SSC, MCYN, KX, RH, LJ, XL, YSC, TK, EST, EZ, MIM, KLH, LB, BT, WYS, YB, JCNC and WJ. Data were acquired by RCWM, CH, VKLL, JW, YW (Shanghai), YJK, HF, HML, FJ, XM, XH, ST, JL (Shanghai), XZ, GX, PK, TFL, GNT, KH, CW, JW, WY, DP-KN, Y-JF, ACWN, HF, APSK, TY, SK-WT, JW, PCL, NLST, JL (Singapore), T-YW, J-YL, SM, SSC, MCYN, KX, APM, SK, RH, LJ, XL, YSC, TK, EST, EZ, MIM, KLH, LB, BT, WYS, YB, JCNC and WJ. Statistical analysis was performed by CHT, RZ, XS, JSKH, YW (Hong Kong) and CH. The data were analysed by CHT, RZ, RCWM, CH, MJG, XS, KH, JSKH, CW, HL, LL, YW (Shanghai), KP, AGDW, JWL, TY, T-FC, SKWT, HYS, APM, SK, EZ, WYS, JCNC and WJ. The manuscript was written by RCWM, CH, CHT, RZ, JWL, TFL, WYS, JCNC and WJ. All authors contributed to the drafting or critical revision of the manuscript for important intellectual content. All authors reviewed and approved the final manuscript.

References

1. Shaw JE, Sicree RA, Zimmet PZ (2010) Global estimates of the prevalence of diabetes for 2010 and 2030. Diabetes Res Clin Pract 87:4–14
2. Chan JC, Malik V, Jia W et al (2009) Diabetes in Asia: epidemiology, risk factors, and pathophysiology. JAMA 301:2129–2140
3. Ramachandran A, Ma RC, Snehalatha C (2010) Diabetes in Asia. Lancet 375:408–418
4. Yang W, Lu J, Weng J et al (2010) Prevalence of diabetes among men and women in China. N Engl J Med 362:1090–1101
5. Zeggini E, Scott LJ, Saxena R et al (2008) Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat Genet 40:638–645
6. 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
7. Hu C, Wang C, Zhang R et al (2009) Variations in KCNQ1 are associated with type 2 diabetes and beta cell function in a Chinese population. Diabetologia 52:1322–1325
8. Hu C, Zhang R, Wang C et al (2010) Variants from GIPR, TCF7L2, DGKB, MADD, CRY2, GLIS3, PROX1, SLC30A8 and IGF1 are associated with glucose metabolism in the Chinese. PLoS One 5:e15542
9. Hu C, Zhang R, Wang C et al (2009) PPARG, KCNJ11, CDKAL1, CDKN2A-CDKN2B, IDE-KIF11-HHEX, IGF2BP2 and SLC30A8 are associated with type 2 diabetes in a Chinese population. PLoS One 4:e7643
10. 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
11. Tam CH, Ho JS, Wang Y et al (2010) Common polymorphisms in MTNR1B, G6PC2 and GCK are associated with increased fasting plasma glucose and impaired beta-cell function in Chinese subjects. PLoS One 5:e11428
12. Yasuda K, Miyake K, Horikawa Y et al (2008) Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. Nat Genet 40:1092–1097
13. Unoki H, Takahashi A, Kawaguchi T et al (2008) SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations. Nat Genet 40:1098–1102
14. Kooner JS, Saleheen D, Sim X et al (2011) Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. Nat Genet 43:984–989
15. Cho YS, Chen CH, Hu C et al (2011) Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. Nat Genet 44:67–72
16. Cui B, Zhu X, Xu M et al (2010) A genome-wide association study confirms previously reported loci for type 2 diabetes in Han Chinese. PLoS One 6:e22353
17. Tsai FJ, Yang CF, Chen CC et al (2010) A genome-wide association study identifies susceptibility variants for type 2 diabetes in Han Chinese. PLoS Genet 6:e1000847
18. Shu XO, Long J, Cai Q et al (2010) Identification of new genetic risk variants for type 2 diabetes. PLoS Genet 6:e1001127
19. 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
20. 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
21. Li H, Gan W, Lu L et al (2013) A genome-wide association study identifies GRK5 and RASGRP1 as type 2 diabetes loci in Chinese Hans. Diabetes 62:291–298
22. Purcell S, Neale B, Todd-Brown K et al (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis. Am J Hum Genet 81:559–575
23. Li Y, Willer C, Sanna S, Abecasis G (2009) Genotype imputation. Annu Rev Genom Hum Genet 10:387–406
24. Magi R, Morris AP (2010) GWAMA: software for genome-wide association meta-analysis. BMC Bioinforma 11:288
25. DerSimonian R, Laird N (1986) Meta-analysis in clinical trials. Control Clin Trials 7:177–188
26. Gaulton KJ, Nammo T, Pasquali L et al (2010) A map of open chromatin in human pancreatic islets. Nat Genet 42:255–259
27. Stitzel ML, Sethupathy P, Pearson DS et al (2010) Global epigenomic analysis of primary human pancreatic islets provides insights into type 2 diabetes susceptibility loci. Cell Metab 12:443–455
28. 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
29. 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

30. Kim YA, Wuchty S, Przytycka TM (2011) Identifying causal genes and dysregulated pathways in complex diseases. PLoS Comput Biol 7:e1001095

31. Barrenas F, Chavali S, Alves AC et al (2012) Highly interconnected genes in disease-specific networks are enriched for disease-associated polymorphisms. Genome Biol 13:R46

32. Lee I, Blom UM, Wang PI, Shim JE, Marcotte EM (2011) Prioritizing candidate disease genes by network-based boosting of genome-wide association data. Genome Res 21:1109–1121

33. Ong RT, Teo YY (2010) varLD: a program for quantifying variation in linkage disequilibrium patterns between populations. Bioinformatics 26(9):1269–1270

34. Brun T, Franklin I, St-Onge L et al (2004) The diabetes-linked transcription factor PAX4 promotes {beta}-cell proliferation and survival in rat and human islets. J Cell Biol 167:1123–1135

35. Li Y, Nagai H, Ohno T et al (2006) Aberrant DNA demethylation in promoter region and aberrant expression of mRNA of PAX4 gene in hematologic malignancies. Leuk Res 30:1547–1553

36. Habener JF, Kemp DM, Thomas MK (2005) Minireview: transcriptional regulation in pancreatic development. Endocrinology 146:1025–1034

37. Wang J, Elghazi L, Parker SE et al (2004) The concerted activities of Pax4 and Nkx2.2 are essential to initiate pancreatic beta-cell differentiation. Dev Biol 266:178–189

38. Mauvais-Jarvis F, Smith SB, Le May C et al (2004) PAX4 gene variations predispose to ketosis-prone diabetes. Hum Mol Genet 13:3151–3159

39. Shimajiri Y, Sanke T, Furuta H et al (2001) A missense mutation of Pax4 gene (R121W) is associated with type 2 diabetes in Japanese. Diabetes 50:2864–2869

40. Tokuyama Y, Matsui K, Ishizuka T, Egashira T, Kanatsuka A (2006) The Arg121Trp variant in PAX4 gene is associated with beta-cell dysfunction in Japanese subjects with type 2 diabetes mellitus. Metabolism 55:213–216

41. Plengvidhya N, Kooptiwut S, Songtawee N et al (2007) PAX4 mutations in Thais with maturity onset diabetes of the young. J Clin Endocrinol Metab 92:2821–2826

42. McCarthy MI (2010) Genomics, type 2 diabetes, and obesity. N Engl J Med 363:2339–2350

43. Silander K, Mohlke KL, Scott LJ et al (2004) Genetic variation near the hepatocyte nuclear factor-4 alpha gene predicts susceptibility to type 2 diabetes. Diabetes 53:1141–1149

44. Bonnefond A, Clement N, Fawcett K et al (2012) Rare MTNR1B variants impairing melatonin receptor 1B function contribute to type 2 diabetes. Nat Genet 44:297–301

45. Luke MR, Houghton F, Perugini MA, Gleeson PA (2005) The trans-Golgi network GRIP-domain proteins form alpha-helical homodimers. Biochem J 388:835–841

46. Lebeda RA, Haun RS (1999) Cloning and characterization of the human ADP-ribosylation factor 4 gene. Gene 237:209–214

47. Song F, Srinivasan M, Aalinkeel R, Patel MS (2001) Use of a cDNA array for the identification of genes induced in islets of suckling rats by a high-carbohydrate nutritional intervention. Diabetes 50:2053–2060

48. Broadhurst MK, Lee RS, Hawkins S, Wheeler TT (2005) The p100 EBNA-2 coactivator: a highly conserved protein found in a range of exocrine and endocrine cells and tissues in cattle. Biochim Biophys Acta 1681:126–133

49. Dupont S, Vionnet N, Chevre JC et al (1999) No evidence of linkage or diabetes-associated mutations in the transcription factors BETA2/NEUROD1 and PAX4 in type II diabetes in France. Diabetologia 42:480–484

50. Dusatkova P, Vesela K, Pruhova S, Lebl J, Cinek O (2010) Lack of PAX4 mutations in 53 Czech MODYX families. Diabet Med 27:1459–1460