

ORIGINAL ARTICLE

Genetic variants on chromosome 6p21.1 and 6p22.3 are associated with type 2 diabetes risk: a case–control study in Han Chinese

Feng Lu^{1,5}, Yun Qian^{1,2,5}, Huizhang Li¹, Meihua Dong², Yudi Lin², Jiangbo Du¹, Yuan Lin¹, Jian Chen², Chong Shen¹, Guangfu Jin¹, Juncheng Dai¹, Zhibin Hu^{1,3,4} and Hongbing Shen^{1,3,4}

Recent genome-wide association studies have identified several single nucleotide polymorphisms (SNPs) on chromosome 6p21.1 and 6p22.3 as type 2 diabetes (T2D) susceptibility loci in the European and Japanese populations. However, these SNPs have not been well evaluated in Chinese population. Here, we performed a case–control study with 2925 T2D cases and 3281 controls in a Chinese population. We used TaqMan OpenArray and Sequenom MassARRAY to genotype the four SNPs (rs4712523, rs7756992, rs4712524 and rs6931514) in *CDKAL1* (cyclin-dependent kinase 5 regulatory subunit-associated protein 1-like 1) at 6p22.3 and one SNP (rs9472138) near vascular endothelial growth factor A (*VEGFA*) at 6p21.1. All the five SNPs were significantly associated with T2D risk with overall effects (odds ratio, OR) from 1.19 to 1.29 in the additive genetic model (rs6931514: OR=1.29, 95% confidence intervals (95% CI)=1.19–1.39, $P=5.6 \times 10^{-10}$; rs7756992: OR=1.23, 95% CI=1.15–1.32, $P=1.2 \times 10^{-8}$; rs4712523: OR=1.25, 95% CI=1.15–1.35, $P=3.8 \times 10^{-8}$; rs4712524: OR=1.24, 95% CI=1.15–1.35, $P=6.8 \times 10^{-8}$; rs9472138: OR=1.19, 95% CI=1.05–1.34, $P=0.06$). Conditional analysis identified two independent signals (rs6931514 at 6p22.3 and rs9472138 at 6p21.1) that were significantly associated with T2D. Compared with the wild homozygote of rs6931514 and rs9472138, subjects with variant alleles of the two SNPs had increased risk for T2D susceptibility in a dose-response manner ($P_{\text{trend}}=7.4 \times 10^{-12}$). Our findings indicated that genetic variants of *CDKAL1* and *VEGFA* on chromosome 6 may contribute to T2D risk in Chinese population, especially for rs9472138 at 6p21.1 identified for the first time to significantly increase the T2D risk in Chinese individuals.

Journal of Human Genetics (2012) 57, 320–325; doi:10.1038/jhg.2012.25; published online 22 March 2012

Keywords: *CDKAL1*; genome-wide association study; polymorphism; type 2 diabetes; *VEGFA*

INTRODUCTION

Type 2 diabetes (T2D) has become one of the fastest growing public health problems worldwide. T2D affects at least 6% of the world's population and the prevalence is expected to double by 2025, with the greatest increases in the developing countries.^{1,2} In China, it was estimated that approximately 92.4 million adults were suffering from T2D per year during 2007–2008 and there was an increased prevalence in young and middle-aged population.³ T2D is a complex polygenic disorder characterized by insulin resistance and deficient pancreatic beta-cell function.⁴ Although environmental factors have clearly contributed to the rise in prevalence of T2D, genetic factors have an important impact on T2D susceptibility.⁵

Over the past several years, genome-wide association studies (GWAS) have provided tens of new susceptibility loci for T2D,⁶ including a cluster of loci on chromosome 6.^{7–14} Four Caucasian

GWAS had consistently identified *CDKAL1* (6p22.3) as a novel T2D susceptibility gene in 2007. Intronic variants (rs10946398, rs7756992, rs7754840 and rs9465871) in *CDKAL1* (cyclin-dependent kinase 5 (CDK5) regulatory subunit-associated protein 1-like 1) had significant associations with the risk of T2D.^{7–10} Subsequently, four new single nucleotide polymorphisms (SNPs) (rs4712524, rs4712523, rs6931514 and rs10440833) in *CDKAL1* were also reported to increase the risk of T2D in GWAS of the Japanese and Caucasian population.^{11–14} Simultaneously, Zeggini *et al.*^{7,11} detected another two new T2D susceptible signals at 6p21.1 (the maker SNP: rs9472138 and rs9369425) neighboring to vascular endothelial growth factor A (*VEGFA*) in a Caucasian T2D GWAS and a meta-analysis of T2D GWAS. However, the two SNPs were failed to replicate in the final validations: further studies will be required to establish the associations with the increased T2D risk.

¹Department of Epidemiology and Biostatistics, MOE Key Laboratory of Modern Toxicology, School of Public Health, Cancer Center, Nanjing Medical University, Nanjing, China; ²Department of Chronic Non-Communicable Disease Control, Wuxi Center for Disease Control and Prevention, Wuxi, China; ³Section of Clinical Epidemiology, Jiangsu Key Lab of Cancer Biomarkers, Prevention and Treatment, Cancer Center, Nanjing Medical University, Nanjing, China and ⁴State Key Laboratory of Reproductive Medicine, Nanjing Medical University, Nanjing, China. ⁵These authors contributed equally to this work

Correspondence: Dr H Shen, Department of Epidemiology and Biostatistics, MOE Key Laboratory of Modern Toxicology, School of Public Health, Cancer Center, Nanjing Medical University, 140 Hanzhong Road, Nanjing, Jiangsu 210029, China.

E-mail: hbshen@njmu.edu.cn

Received 25 November 2011; revised 30 January 2012; accepted 1 February 2012; published online 22 March 2012

CDKAL1 resides at chromosome 6p22.3 and encodes a 579-residue, 65 kD protein with unclear molecular function. However, *CDKAL1* shares considerable protein domain and amino-acid homology with *CDK5* regulatory subunit-associated protein 1 (*CDK5RAP1*), a known inhibitor of *CDK5* activation.¹⁵ *CDK5* has been implicated in the regulation of pancreatic beta cell function through the formation of *p35/CDK5* complexes that downregulate insulin expression.¹⁶ *VEGFA*, locates at the downstream of *CDKAL1* at the same chromosome (6p21.1). As a major member of *VEGF* family, *VEGFA* is a mitogen that specifically acts on endothelial cells and has various effects, including mediating increased vascular permeability, inducing angiogenesis, cell growth, migration and inhibition of apoptosis.¹⁷ It has been recognized that T2D mainly accompanies with the increased expression of *VEGF* in numerous tissues (retina and glomeruli) as a response to hyperglycemia and tissue ischemia, finally leading to complications such as retinopathy and diabetic nephropathy.¹⁸ All these findings further support the biological plausibility that genetic variants of *CDKAL1* and *VEGFA* may affect the risk of T2D.

Up to now, most of the GWAS were conducted among populations of non-Chinese ancestry and the contributions of these variants to the T2D risk in Han Chinese are poorly characterized. Nevertheless, some most significant SNPs on chromosome 6 have been successfully replicated in many case-control studies of Chinese population but no study has conducted to include all the SNPs identified by GWAS on chromosome 6 associated with T2D risk. For instance, the association of SNP rs9472138 neighboring to *VEGFA* with the risk of T2D has not been confirmed in Chinese after GWAS exploration and the minor allele frequency of this variant is lower in Chinese (0.116), Japanese (0.081) and Africans (0.137) than in Caucasians (0.239), suggesting that there may be a heterogeneity of the associations with T2D among different populations. Here, we performed a case-control study including 2925 T2D cases and 3281 controls in a Han Chinese population to further determine the associations between these SNPs on chromosome 6p21.1 and 6p22.3 and T2D risk.

MATERIALS AND METHODS

Study subjects

We performed a case-control study of T2D in a Chinese population and our samples were composed of two parts. One part of samples (1200 T2D cases and 1200 orthoglycemic controls) were recruited from a cross-sectional survey of more than 21 000 participants aged over 30 years in a community-based non-communicable diseases screening program conducted in Wuxi City, Jiangsu Province of southeastern China, from April to July 2007, and the other part of 1725 cases and 2081 controls were selected from an ongoing, population-based cohort study of about 40 000 subjects in Changzhou and Nantong cities in Jiangsu Province during 2004 and 2008. All these cases and controls were unrelated ethnic Han Chinese. The cases were defined by 2003 ADA (the American Diabetes Association) criteria (fasting blood glucose, $\text{FBG} \geq 7.0 \text{ mmol l}^{-1}$) or previously diagnosed T2D. All the controls were randomly selected from the subjects with normal fasting glucose defined as $\text{FBG} < 5.6 \text{ mmol l}^{-1}$ and were frequency-matched to cases on the age of interview, sex and residential area (urban and rural).^{19,20} The subjects who self-reported having diabetes, hypertension, coronary heart disease and cancer in the baseline or follow-up were all excluded from the controls. A questionnaire was completed by trained interviewers to collect information on demographic characteristics, risk factors, physical activity and disease history after the written informed consent was obtained from every subject. At the same time, physical examinations including height, weight and blood pressure were conducted and the average level of twice of blood pressure test was used. BMI was calculated as weight (in kg)/height (in m^2) and overweight or obesity was defined as $\text{BMI} \geq 24 \text{ kg m}^{-2}$.²¹ Approximately 5 ml blood was collected from each participant early in the morning after an overnight fast. FBG, triglycerides, cholesterol and high-density lipoprotein cholesterol were measured enzymatically

by Hitachi 7180 Biochemistry Auto-analyzer (Hitachi High-Tech Science Systems Corporation, Hitachinaka, Japan) following the manufacture's instructions.

As a result, a total of 2925 cases and 3281 controls who completed the interviews, physical examinations and donated the blood samples with biochemical measurements were included in the final analysis. This study was approved by the Institutional Review Board of Nanjing Medical University (Nanjing, China).

SNPs selection and genotyping

Based on the recent GWA studies of T2D, we identified eight SNPs in *CDKAL1* (6p22.3), showing associations with the risk of T2D, including rs4712523, rs10946398, rs7754840, rs7756992, rs9465871, rs10440833, rs4712524 and rs6931514. The minor allele frequencies of all these eight SNPs were higher than 0.05 in Chinese Han population (CHB). Based on the HapMap database (phase II, 8 November, on NCBI B36 assembly, dbSNP b126), we selected two representative SNPs (rs4712523 and rs7756992) with $r^2 < 0.8$ among the first six SNPs in CHB (Figure 1). The last two SNPs (rs4712524 and rs6931514) that had no linkage disequilibrium (LD) information in HapMap were also selected for genotyping. In addition, SNPs rs9472138 and rs9369425 close to *VEGFA* (6p21.1) significantly associated with T2D risk were also in complete LD ($r^2=1.0$), and we selected rs9472138 in the validation for its stronger reported association than rs9369425. Finally, five SNPs (rs4712523, rs7756992, rs4712524, rs6931514 and rs9472138) were included in the present study.

Genomic DNA was extracted from peripheral blood leukocytes by proteinase K digestion and followed by phenol/chloroform-based method. Two genotyping platforms were used to genotype the samples. For the first part of the 1200 T2D cases and 1200 controls, we used TaqMan OpenArray Genotyping System (Life Technologies, Carlsbad, CA, USA), a medium-throughput genotyping platform. Normalized DNA samples were loaded and amplified on 48-sample arrays following the manufacturer's protocol. Equal numbers of cases and controls were assayed and two no template controls in each chip were used for quality control. The overall call rates were ranged from 98.5 to 99.3% for this assay. For the other part of the 1725 cases and 2081 controls, all five SNPs were genotyped using the iPLEX Sequenom MassARRAY platform (Sequenom, Inc., San Diego, CA, USA). Genotyping was performed without knowing the subjects' case or control status and two no template controls in each 384-sample array chip were used for quality control. The genotype call rates for the Sequenom assay were all higher than 99.3%. Randomly selected 96 duplicated samples were assessed by the two platforms and the accordance achieved was 100%.

Statistical analyses

Student's *t*-test (for continuous variables) or χ^2 -test (for categorical variables) were used to analyze the differences in selected demographic variables, risk factors and frequencies of variant genotypes between the cases and controls. Hardy-Weinberg equilibrium was tested by a goodness-of-fit χ^2 -test to compare the observed genotype frequencies with the expected ones among the controls. LD between SNPs was evaluated using Haploview version 4.1 (Broad Institute, Cambridge, MA, USA). Odds ratios (ORs) and 95% confidence intervals (CIs) were computed by using multivariate logistic regression with adjustment for age, sex and BMI to estimate the associations between each SNP and the risk of T2D. Conditional regression analyses on each SNP were conducted by using logistic regression with adjustment for age, sex, BMI and any of the other four SNPs in additive genetic model.^{11,22} The haplotypes of genotyped SNPs were inferred by PHASE 2.1. Between-group heterogeneity test in a stratified analysis was assessed by the χ^2 -based *Q* test. We used PS 3.0.2 software (Dupont WD and Plummer WD, Nashville, TN, USA) to perform the power estimation. All the statistical analyses were performed by the PLINK 1.07 (Purcell S, Cambridge, MA, USA) and STATA 11.1 software (StataCorp, College Station, TX, USA). $P < 0.05$ was the criterion of statistical significance, and all statistical tests were two sided.

RESULTS

The distributions of selected characteristics of the 2925 T2D cases and 3281 controls are shown in Table 1. There were no significant differences in the variables of sex, smoking and drinking status between the cases and the controls ($P=0.958$, 0.556 and 0.336, respectively). However, the mean age of the cases (58.21 ± 10.11) was higher than that of the controls (56.57 ± 9.88). As expected,

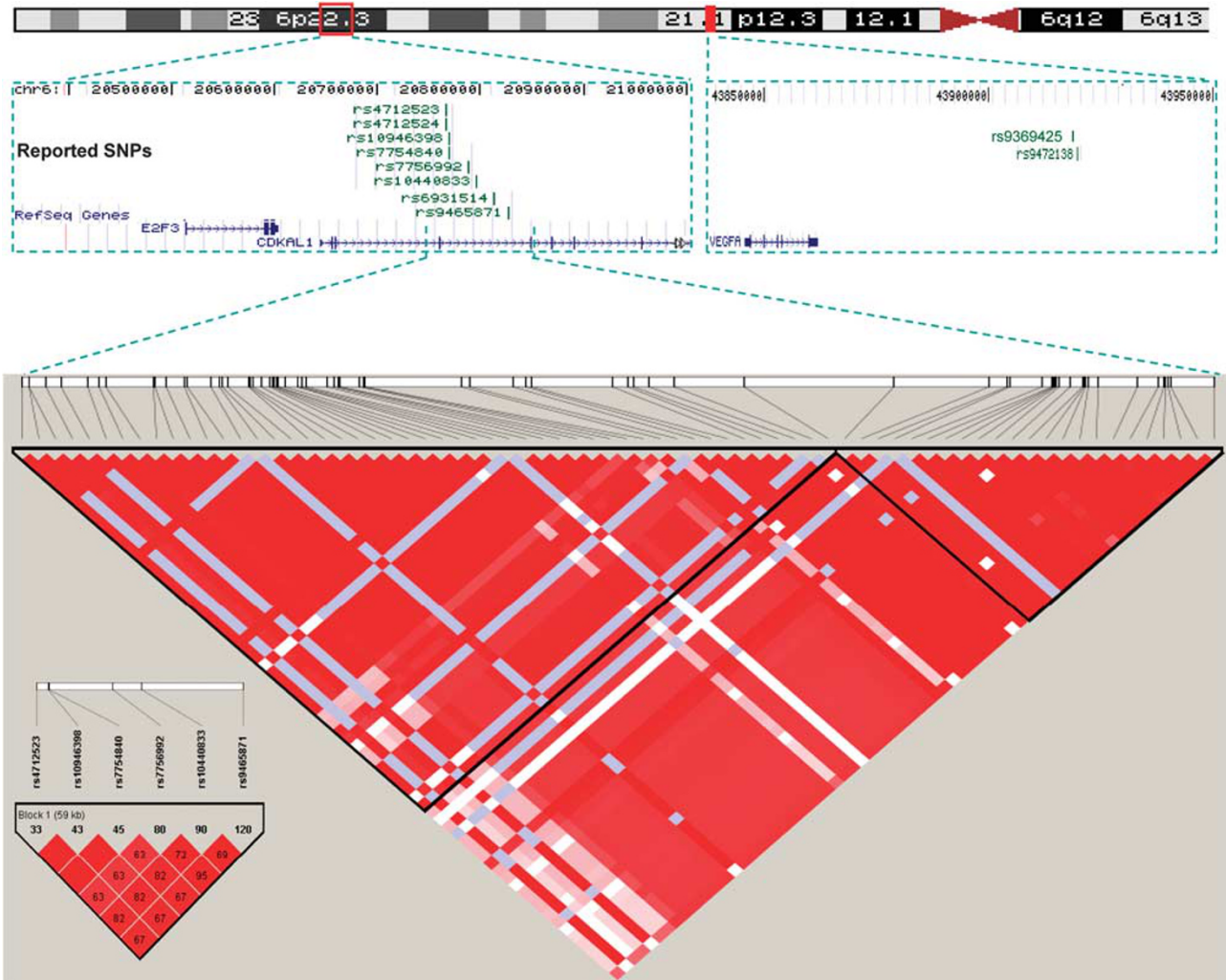


Figure 1 Overview of the GWAS reported loci on 6p22.3 and 6p21.1 associated with T2D risk. The upper panel shows a view of the genomic region in part of 6p22.3 (20 400 000–21 000 000) and 6p21.1 (43 840 000–43 950 000) from the UCSC browser Build 36 assembly (hg18), marking with nine SNPs, which were reported to associate with the risk of T2D in GWAS. The lower panel shows the LD plot (20 751 201–20 836 492) of intron 5 in *CDKAL1* gene of Chinese descent (the block was defined by the confidence intervals). The lower left thumbnail shows the r^2 value between six reported SNPs in *CDKAL1* (rs6931514 and rs4712524 that have no LD information in Hapmap database are directly included in our study).

compared with the control subjects, the T2D cases had a significantly higher BMI, FBG, triglycerides, cholesterol, SBP, DBP and lower high-density lipoprotein cholesterol (all P -values < 0.0001).

The genotype distributions of the five selected variants between the T2D cases and controls are shown in Table 2. The observed genotype frequencies for five SNPs were all consistent with the Hardy–Weinberg equilibrium among the controls ($P=0.53, 0.58, 0.83, 0.54$ and 0.06 for rs6931514, rs7756992, rs4712523, rs4712524, and rs9472138, respectively). Considering multiple comparisons, we used Bonferroni correction on the results of additive genetic model and $P=0.01(0.05/5)$ was considered as significance threshold. Multivariate logistic regression analyses showed that all the five SNPs were significantly associated with increased T2D risk with overall effects (OR) from 1.19 to 1.29 (rs6931514: adjusted OR=1.29, 95% CI=1.19–1.39, $P=5.6 \times 10^{-10}$; rs7756992: adjusted OR=1.23, 95% CI=1.15–1.32, $P=1.2 \times 10^{-8}$; rs4712523: adjusted OR=1.25, 95% CI=1.15–1.35, $P=3.8 \times 10^{-8}$; rs4712524: adjusted OR=1.24, 95% CI=1.15–1.35, $P=6.8 \times 10^{-8}$; rs9472138: adjusted OR=1.19, 95% CI=1.05–1.34, $P=0.006$).

We then used conditional analysis to test the independence of these SNPs. As shown in Table 3, after controlling any of the other four SNPs, rs9472138 showed substantially significant association with T2D risk, while rs6931514 and rs7756992 were not significant after adjustment with each other, as the two SNPs were in highly LD ($D'/r^2=0.99/0.97$). Moreover, the effect of rs4712523 and rs4712524 on T2D risk was abolished after condition on the other three SNPs except rs9472138. Therefore, only two SNPs were considered independent signals of T2D susceptibility on 6p22.3 (rs6931514) and 6p21.1 (rs9172138).

Based on the results of the conditional analysis, we included SNPs rs9472138 and rs6931514 with independent effect in further combined analysis. The combined effect was calculated by summing up the number of risk alleles for each subject who had the genotyping information of these two SNPs. As shown in Table 4, compared with the group of '0' risk allele, we found significantly increased risk for T2D associated with the combined genotypes containing 'two' risk alleles (adjusted OR=1.59, 95% CI=1.35–1.88, $P=5.6 \times 10^{-8}$) and 'three to four' risk alleles (adjusted OR=1.86, 95%

Table 1 Distributions of selected variables between type 2 diabetes cases and controls

Variables	Cases N=2925(%)	Controls N=3281(%)	P ^a
Age, year ^b	58.21 ± 10.11	56.57 ± 9.88	<0.0001
≤ 56	1138 (38.9)	1451 (44.2)	
> 56	1786 (61.1)	1830 (55.8)	
Sex			0.958
Females	1827 (62.5)	2047 (62.4)	
Males	1098 (37.5)	1234 (37.6)	
Smoking status ^b			0.556
No	2185 (75.3)	2431 (74.6)	
Yes	717 (24.7)	827 (25.4)	
Drinking status ^b			0.336
No	2345 (81.5)	2619 (80.5)	
Yes	534 (18.5)	635 (19.5)	
BMI, kg m ^{-2b}	25.05 ± 3.50	22.12 ± 2.63	<0.0001
< 24	1124 (38.7)	2620 (80.0)	<0.0001
≥ 24	1778 (61.3)	655 (20.0)	
Fasting blood glucose, mmol l ^{-1b}	9.03 ± 3.52	4.52 ± 0.53	<0.0001
Triglycerides, mmol l ^{-1b}	2.57 ± 2.47	1.09 ± 0.45	<0.0001
Cholesterol, mmol l ^{-1b}	4.98 ± 1.44	4.39 ± 0.80	<0.0001
HDL, mmol l ^{-1b}	1.46 ± 0.48	1.62 ± 0.38	<0.0001
Systolic blood pressure, mm Hg ^b	135.47 ± 20.99	117.68 ± 14.17	<0.0001
Diastolic blood pressure, mm Hg ^b	80.36 ± 10.87	74.14 ± 9.24	<0.0001

Abbreviations: BMI, body mass index; HDL, high-density lipoprotein cholesterol.

^at-tests and χ^2 -tests were used for continuous or categorical variables, respectively, in combined population.^bIn all, 1 subject was absent for information on age; 46 subjects were absent for information on smoking and 73 subjects were absent for information on drinking; 29, 24, 21, 21, 21, 7 and 8 subjects were absent for information on BMI, fasting blood glucose, triglycerides, cholesterol, HDL, systolic blood pressure and diastolic blood pressure, respectively. Data are means ± s.d. or numbers (percentage).CI=1.44–2.38, $P=1.4 \times 10^{-6}$) in a dose-response effect (P for trend= 7.4×10^{-12}).

As four of the five SNPs were located in one LD block within *CDKAL1* gene, the pattern of risk association was further evaluated by examination of the common haplotypes derived from observed genotypes (Supplementary Table 1). Compared with the most common haplotype AAAA, the GGGG haplotype that was consisted of risk alleles of four SNPs was associated with a 33% elevated risk of T2D (adjusted OR=1.33, 95% CI=1.22–1.45, $P=1.5 \times 10^{-10}$). Meanwhile, haplotype GGAA was also significantly associated with a 1.17-fold increased T2D risk (adjusted OR=1.17, 95% CI=1.03–1.33, $P=0.015$).

We then conducted the stratification analysis for two independent SNPs rs6931514 and rs9472138 and the combined genotypes of the two SNPs by age, sex, BMI, smoking and drinking status. As shown in Supplementary Table 2, heterogeneity test showed no significant heterogeneity in between any two strata.

DISCUSSION

In this case-control study of T2D, we investigated the associations between four SNPs in *CDKAL1* (rs6931514, rs7756992, rs4712523 and rs4712524) and one SNP neighboring *VEGFA* (rs9472138) on chro-

Table 2 Genotype distributions of the five SNPs at chromosome 6 between T2D cases and controls and their associations with T2D risk

Genotype	Cases N=2925(%)	Controls N=3281(%)	OR (95% CI) ^a	P ^a
<i>rs6931514</i>				
AA	586 (20.19)	814 (25.02)	1.00	
AG	1415 (48.76)	1608 (49.43)	1.20 (1.04–1.38)	0.014
GG	901 (31.05)	831 (25.55)	1.64 (1.40–1.92)	1.1×10^{-9}
Additive			1.29 (1.19–1.39)	5.6×10^{-10}
<i>rs7756992</i>				
AA	570 (19.63)	802 (24.57)	1.00	
AG	1423 (49.02)	1615 (49.48)	1.21 (1.04–1.39)	0.011
GG	910 (31.35)	847 (25.95)	1.65 (1.41–1.94)	6.6×10^{-10}
Additive			1.23 (1.15–1.32)	1.2×10^{-8}
<i>rs4712523</i>				
AA	848 (29.27)	1120 (34.37)	1.00	
AG	1409 (48.64)	1576 (48.36)	1.15 (1.01–1.31)	0.031
GG	640 (22.09)	563 (17.28)	1.60 (1.36–1.87)	1.3×10^{-8}
Additive			1.25 (1.15–1.35)	3.8×10^{-8}
<i>rs4712524</i>				
AA	880 (30.36)	1156 (35.46)	1.00	
AG	1397 (48.19)	1556 (47.73)	1.14 (1.01–1.30)	0.040
GG	622 (21.46)	548 (16.81)	1.59 (1.35–1.87)	2.0×10^{-8}
Additive			1.24 (1.15–1.35)	6.8×10^{-8}
<i>rs9472138</i>				
CC	2230 (76.82)	2598 (79.57)	1.00	
CT	628 (21.63)	617 (18.90)	1.24 (1.08–1.43)	0.002
TT	45 (1.55)	50 (1.53)	1.10 (0.70–1.72)	0.674
Additive			1.19 (1.05–1.34)	0.006

^aOR (95%CI) and P -value were calculated in logistic regression with adjustment for age, sex and BMI.

some 6 and risk of T2D in a Chinese population. We found that all these five SNPs could significantly increase T2D risk on the basis of community sources of 2925 T2D cases and 3281 controls.

Among the four replicated SNPs in the *CDKAL1* gene, rs7756992 had shown reproducible association with the risk of T2D in both European population and Chinese populations.^{8,23,24} Hu *et al.*²³ reported a significant association between rs7756992 and risk of T2D in 1849 cases and 1785 controls from Shanghai, which was consistent to our study. Furthermore, some replication studies found that rs7756992 was significantly associated with T2D-related traits (for example, FBG, A1C, HOMA-B and corrected insulin response), indicating that this variant may confer risk of T2D through impaired function of beta cells or reduced insulin secretion.^{8,24} However, this hypothesis needs to be explored by biological assays in future studies.

Compared with the above well-replicated SNPs, the other three SNPs at 6p22.3 have rarely replicated after GWAS discovery, and have not reported in Chinese populations. rs4712523 and rs4712524 were firstly identified to be associated with T2D risk in Japanese GWAS.^{12,13} Similar to the results by Japanese, rs4712523 and rs4712524 GG genotypes were significantly associated with increased risk of T2D in current study. Based on HapMap CHB database, rs4712523 was in very high LD with rs10946398, rs7754840 and rs10440833 ($r^2=1.0$, 1.0 and 0.82, respectively), which all showed associations with T2D risk both in European GWAS and in Chinese replication studies.^{19,25–28}

Table 3 Conditional regression analysis and LD values between the 5 SNPs

Primary SNP	Risk allele	Unadjusted P ^a	Adjusted for SNPs	Adjusted P ^b	D'/r ²
rs6931514	G	5.6×10 ⁻¹⁰	rs7756992	0.449	0.99/0.97
			rs4712523	0.001	0.81/0.46
			rs4712524	0.001	0.84/0.48
			rs9472138	4.7×10 ⁻¹⁰	0.01/0.00
rs7756992	G	1.2×10 ⁻⁸	rs6931514	0.079	—
			rs4712523	3.9×10 ⁻⁴	0.82/0.46
			rs4712524	3.1×10 ⁻⁴	0.85/0.48
			rs9472138	3.2×10 ⁻¹⁰	0.00/0.00
rs4712523	G	3.8×10 ⁻⁸	rs6931514	0.083	—
			rs7756992	0.140	—
			rs4712524	0.285	0.99/0.96
			rs9472138	3.6×10 ⁻⁸	0.05/0.00
rs4712524	G	6.8×10 ⁻⁸	rs6931514	0.155	—
			rs7756992	0.234	—
			rs4712523	0.959	—
			rs9472138	6.3×10 ⁻⁸	0.03/0.00
rs9472138	T	0.006	rs6931514	0.003	—
			rs7756992	0.006	—
			rs4712523	0.005	—
			rs4712524	0.005	—

Abbreviations: LD, linkage disequilibrium; SNP, single nucleotide polymorphism.
^aP-value calculated in additive genetic model by logistic regression with adjustment for age, sex and BMI.
^bP-value for the primary SNP calculated in additive genetic model by logistic regression with adjustment for age, sex, BMI and one additional SNP in column 'Adjusted for SNPs'.

Table 4 Cumulative effects of rs6931514 and rs9472138 on T2D risk

The number of alleles ^a	Cases	Controls	OR (95% CI) ^b	P ^b
0	455 (15.7)	637 (19.7)	1	
1	1193 (41.3)	1457 (44.9)	1.14 (0.97–1.34)	0.109
2	1015 (35.1)	951 (29.3)	1.59 (1.35–1.88)	5.6×10 ⁻⁸
3–4	228 (7.89)	197 (6.1)	1.86 (1.44–2.38)	1.4×10 ⁻⁶
P for trend ^b =7.4×10 ⁻¹²				
0	455 (15.7)	637 (19.7)	1	
1–4	2436 (84.3)	2605 (80.3)	1.35 (1.17–1.57)	6.6×10 ⁻⁵

^aThe risk alleles were defined as the number of variant alleles a person carries in two SNPs.
^bAdjusted by age, sex and BMI.

rs6931514 was in high LD with rs7756992 (D'/r²=0.99/0.97), although it was not reported in HapMap CHB database.

Overall, the four associated SNPs all mapped to the same intron (intron 5: 90 kb) of the *CDKAL1* gene on 6p22.3. These variants reside in a large LD block of 200 kb that includes exons 1–5 of *CDKAL1* and the minimal promoter region, likely to contain causative variants (Figure 1). The function of the gene product of *CDKAL1* is little known but it shares large protein domain with CDK5RAP1, which inhibits CDK5 activity and expresses in neuronal tissues. CDK5 has been identified to have an important role in beta cell dysfunction and the decrease of insulin gene expression under glucotoxic conditions.²⁹

Thus, *CDKAL1* could have a similar role in the inhibition of the CDK5 in pancreatic beta cells as CDK5RAP1 in neuronal tissue. Beta cell-specific *CDKAL1* knockout mice models suggested that functional loss of *CDKAL1* affects the accuracy of protein translation, causing the synthesis of abnormal insulin, which triggers estrogen receptor stress in beta cells.^{30,31}

The significant association between the SNP rs9472138 (about 73 kb to the transcription start site of *VEGFA*) and T2D was firstly replicated among Chinese population. Zeggini *et al.* first reported in Caucasian that rs9472138 was associated with T2D in GWAS scan and the first stage validation, but achieved a borderline significant association (P=0.095) in the second stage validation. The OR (adjusted OR=1.19, 95% CI=1.05–1.34, P=0.006 in additive model) in our study was slightly higher than that identified in Caucasian GWAS (adjusted OR=1.06, 95% CI=1.06–1.09), which implied some ethnic differences between these studies for distinct risk allele frequencies (0.110 in our controls vs 0.282 in European population).¹¹ Staiger *et al.*³² reported that rs9472138 tended to associate with more than one measure of insulin secretion and insulin sensitivity in 1578 metabolically characterized non-diabetic German individuals. Furthermore, Sharma *et al.* reported that rs9472138 was significantly associated with the expression of *VEGFA* in transformed lymphocytes from 82 Caucasians, and human adipose and muscle expression of *VEGFA* was also correlated with BMI and insulin sensitivity.³³ By similar screening cis-eQTLs in transformed lymphocytes of African-American subjects, Langberg *et al.*³⁴ found that rs9472138 was associated with nearby transcript expressions (MEAL, SLC35B2, NFKBIE), but the association between the variant and T2D was not successfully validated (P=0.128). A recent replication study with 3388 T2D cases and 2856 controls in a Japanese population failed to repeat the association between rs9472138 and T2D (OR=1.06, P=0.41).³⁵ This finding was inconsistent with our results in Chinese population and might also reflect the difference among different populations. On the other hand, because of the lower allele frequency of rs9472138 that leads to the lower statistical power (about 64%), we could not rule out the possibility of false-positive associations, which needs validation from large studies in the future.

Some limitations of this study need to be addressed. First, there may be a spurious finding by population stratification derived from samples comprising two parts. However, we used the χ^2 -based Q statistic to test for the heterogeneity between groups in a stratified analysis and we found that the P-values of heterogeneity tests were 0.50, 0.64, 0.06, 0.11 and 0.15 for rs6931514, rs7756992, rs4712523, rs4712524 and rs9472138, respectively, indicating that it was unlikely that the associations were resulted from population stratification (data not shown). Second, when this article was in preparation, a T2D GWAS study for the East Asian population consisting 25 079 cases and 29 611 controls was published, reporting that six new T2D loci reaching genome-wide significance, including ZFAND3 (the marker SNP: rs9470794) at 6p22.3 and KCNK16 (the marker SNP: rs1535500) at 6p21.1, which further confirmed that these two susceptible regions might have an important role in T2D.³⁶

In conclusion, we found that five SNPs on chromosome 6 were significantly associated with risk of T2D in Chinese population, especially for SNP rs9472138 at 6p21.1 identified for the first time to significantly increase the T2D risk in Chinese individuals. Further functional studies are warranted to explore the potential mechanisms of these SNPs associated with T2D risk.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by the China National Natural Science Foundation Grant 81072379, Ministry of Health Research Program Grant WKJ2010-2-032, Wuxi Science & Technology Research Program Grant CSE01016 and the Priority Academic Program for the Development of Jiangsu Higher Education Institutions (Public Health and Preventive Medicine).

- Narayan, K. M., Boyle, J. P., Geiss, L. S., Saaddine, J. B. & Thompson, T. J. Impact of recent increase in incidence on future diabetes burden: US, 2005-2050. *Diabetes Care* **29**, 2114-2116 (2006).
- Chan, J. C., Malik, V., Jia, W., Kadowaki, T., Yajnik, C. S., Yoon, K. H. *et al*. Diabetes in Asia: epidemiology, risk factors, and pathophysiology. *JAMA* **301**, 2129-2140 (2009).
- Yang, W., Lu, J., Weng, J., Jia, W., Ji, L., Xiao, J. *et al*. Prevalence of diabetes among men and women in China. *N. Engl. J. Med.* **362**, 1090-1101 (2010).
- Stumvoll, M., Goldstein, B. J. & van Haeften, T. W. Type 2 diabetes: principles of pathogenesis and therapy. *Lancet* **365**, 1333-1346 (2005).
- Moore, A. F. & Florez, J. C. Genetic susceptibility to type 2 diabetes and implications for anti-diabetic therapy. *Annu. Rev. Med.* **59**, 95-111 (2008).
- Imamura, M. & Maeda, S. Genetics of type 2 diabetes: the GWAS era and future perspectives [Review]. *Endocr. J.* **58**, 723-739 (2011).
- Zeggini, E., Weedon, M. N., Lindgren, C. M., Frayling, T. M., Elliott, K. S., Lango, H. *et al*. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* **316**, 1336-1341 (2007).
- Steinthorsdottir, V., Thorleifsson, G., Reynisdottir, I., Benediktsson, R., Jonsdottir, T., Walters, G. B. *et al*. A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat. Genet.* **39**, 770-775 (2007).
- Scott, L. J., Mohlke, K. L., Bonnycastle, L. L., Willer, C. J., Li, Y., Duren, W. L. *et al*. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316**, 1341-1345 (2007).
- Consortium T.W.T.C.C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661-678 (2007).
- Zeggini, E., Scott, L. J., Saxena, R., Voight, B. F., Marchini, J. L., Hu, T. *et al*. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat. Genet.* **40**, 638-645 (2008).
- Unoki, H., Takahashi, A., Kawaguchi, T., Hara, K., Horikoshi, M., Andersen, G. *et al*. SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations. *Nat. Genet.* **40**, 1098-1102 (2008).
- Takeuchi, F., Serizawa, M., Yamamoto, K., Fujisawa, T., Nakashima, E., Ohnaka, K. *et al*. Confirmation of multiple risk loci and genetic impacts by a genome-wide association study of type 2 diabetes in the Japanese population. *Diabetes* **58**, 1690-1699 (2009).
- Voight, B. F., Scott, L. J., Steinthorsdottir, V., Morris, A. P., Dina, C., Welch, R. P. *et al*. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet.* **42**, 579-589 (2010).
- Ching, Y. P., Pang, A. S., Lam, W. H., Qi, R. Z. & Wang, J. H. Identification of a neuronal Cdk5 activator-binding protein as Cdk5 inhibitor. *J. Biol. Chem.* **277**, 15237-15240 (2002).
- Wei, F. Y., Nagashima, K., Ohshima, T., Saheki, Y., Lu, Y. F., Matsushita, M. *et al*. Cdk5-dependent regulation of glucose-stimulated insulin secretion. *Nat. Med.* **11**, 1104-1108 (2005).
- Olsson, A. K., Dimberg, A., Kreuger, J. & Claesson-Welsh, L. VEGF receptor signalling in control of vascular function. *Nat. Rev. Mol. Cell Biol.* **7**, 359-371 (2006).
- Takahashi, H. & Shibuya, M. The vascular endothelial growth factor (VEGF)/VEGF receptor system and its role under physiological and pathological conditions. *Clin. Sci. (Lond.)* **109**, 227-241 (2005).
- Ng, M. C., Park, K. S., Oh, B., Tam, C. H., Cho, Y. M., Shin, H. D. *et al*. 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 (2008).
- Liu, Y., Liu, Z., Song, Y., Zhou, D., Zhang, D., Zhao, T. *et al*. Meta-analysis added power to identify variants in FTO associated with type 2 diabetes and obesity in the Asian population. *Obesity (Silver Spring)* **18**, 1619-1624 (2010).
- Liu, C., Ma, X., Ma, Y. T., Liu, F., Yang, Y. N., Huang, D. *et al*. Prevalence on overweight and obesity in Han, Uygur and Hazakh in adults from Xinjiang. *Zhonghua Liu Xing Bing Xue Za Zhi* **31**, 1139-1143 (2010).
- Hu, Z., Wu, C., Shi, Y., Guo, H., Zhao, X., Yin, Z. *et al*. A genome-wide association study identifies two new lung cancer susceptibility loci at 13q12.12 and 22q12.2 in Han Chinese. *Nat. Genet.* **43**, 792-796 (2011).
- Hu, C., Zhang, R., Wang, C., Wang, J., Ma, X., Lu, J. *et al*. 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 (2009).
- Wu, Y., Li, H., Loos, R. J., Yu, Z., Ye, X., Chen, L. *et al*. Common variants in CDKAL1, CDKN2A/B, IGF2BP2, SLC30A8, and HHEX/IDE genes are associated with type 2 diabetes and impaired fasting glucose in a Chinese Han population. *Diabetes* **57**, 2834-2842 (2008).
- Liu, Y., Yu, L., Zhang, D., Chen, Z., Zhou, D. Z., Zhao, T. *et al*. Positive association between variations in CDKAL1 and type 2 diabetes in Han Chinese individuals. *Diabetologia* **51**, 2134-2137 (2008).
- Lin, Y., Li, P., Cai, L., Zhang, B., Tang, X., Zhang, X. *et al*. Association study of genetic variants in eight genes/loci with type 2 diabetes in a Han Chinese population. *BMC Med. Genet.* **11**, 97 (2010).
- Wen, J., Ronn, T., Olsson, A., Yang, Z., Lu, B., Du, Y. *et al*. Investigation of type 2 diabetes risk alleles support CDKN2A/B, CDKAL1, and TCF7L2 as susceptibility genes in a Han Chinese cohort. *PLoS. One.* **5**, e9153 (2010).
- Han, X., Luo, Y., Ren, Q., Zhang, X., Wang, F., Sun, X. *et al*. Implication of genetic variants near SLC30A8, HHEX, CDKAL1, CDKN2A/B, IGF2BP2, FTO, TCF2, KCNQ1, and WFS1 in type 2 diabetes in a Chinese population. *BMC Med. Genet.* **11**, 81 (2010).
- Ubeda, M., Rukstalis, J. M. & Habener, J. F. Inhibition of cyclin-dependent kinase 5 activity protects pancreatic beta cells from glucotoxicity. *J. Biol. Chem.* **281**, 28858-28864 (2006).
- Wei, F. Y., Suzuki, T., Watanabe, S., Kimura, S., Kaitsuka, T., Fujimura, A. *et al*. Deficit of tRNA(Lys) modification by Cdkal1 causes the development of type 2 diabetes in mice. *J. Clin. Invest.* **121**, 3598-3608 (2011).
- Wei, F. Y. & Tomizawa, K. Functional loss of Cdkal1, a novel tRNA modification enzyme, causes the development of type 2 diabetes [Review]. *Endocr. J.* **58**, 819-825 (2011).
- Staiger, H., Machicao, F., Kantartzis, K., Schafer, S. A., Kirchhoff, K., Guthoff, M. *et al*. Novel meta-analysis-derived type 2 diabetes risk loci do not determine prediabetic phenotypes. *PLoS. One.* **3**, e3019 (2008).
- Sharma, N. K., Langberg, K. A., Mondal, A. K., Elbein, S. C. & Das, S. K. Type 2 diabetes (T2D) associated polymorphisms regulate expression of adjacent transcripts in transformed lymphocytes, adipose, and muscle from Caucasian and African-American subjects. *J. Clin. Endocrinol. Metab.* **96**, E394-403 (2011).
- Langberg, K. A., Ma, L., Sharma, N. K., Hanis, C. L., Elbein, S. C., Hasstedt, S. J. *et al*. Single nucleotide polymorphisms in JAZF1 and BCL11A gene are nominally associated with type 2 diabetes in African-American families from the GENNID study. *J. Hum. Genet.* **57**, 57-61 (2011).
- Omori, S., Tanaka, Y., Horikoshi, M., Takahashi, A., Hara, K., Hirose, H. *et al*. Replication study for the association of new meta-analysis-derived risk loci with susceptibility to type 2 diabetes in 6,244 Japanese individuals. *Diabetologia* **52**, 1554-1560 (2009).
- Cho, Y. S., Chen, C. H., Hu, C., Long, J., Hee Ong, R. T., Sim, X. *et al*. Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. *Nat. Genet.* **44**, 67-72 (2011).

Supplementary Information accompanies the paper on Journal of Human Genetics website (<http://www.nature.com/jhg>)