# Nonsynonymous Variants in *PAX4* and *GLP1R* Are Associated With Type 2 Diabetes in an East Asian Population

Soo Heon Kwak,[1] Jeesoo Chae,[2] Seungbok Lee,[2] Sungkyoung Choi,[3] Bo Kyung Koo,[4] Ji Won Yoon,[5] Jin-Ho Park,[6] Belong Cho,[6] Min Kyong Moon,[4,7] Soo Lim,[8] Young Min Cho,[1,7] Sanghoon Moon,[9] Young Jin Kim,[9] Sohee Han,[9] Mi Yeong Hwang,[9] Yoon Shin Cho,[10] Myung-Shik Lee,[11] Hak C. Jang,[7,8] Hyun Min Kang,[12] Taesung Park,[13] Nam H. Cho,[14] Kyunga Kim,[15,16] Jong-Il Kim,[2] and Kyong Soo Park[1,7,17]

We investigated ethnicity-specific exonic variants of type 2 diabetes (T2D) and its related clinical phenotypes in an East Asian population. We performed whole-exome sequencing in 917 T2D case and control subjects, and the findings were validated by exome array genotyping in 3,026 participants. In silico replication was conducted for seven nonsynonymous variants in an additional 13,122 participants. Single-variant and gene-based association tests for T2D were analyzed. A total of 728,838 variants were identified by whole-exome sequencing. Among nonsynonymous variants, *PAX4* Arg192His increased risk of T2D and *GLP1R* Arg131Gln decreased risk of T2D in genome-wide significance (odds ratio [OR] 1.48, $P = 4.47 \times 10^{-16}$ and OR 0.84, $P = 3.55 \times 10^{-8}$, respectively). Another variant at *PAX4* 192 codon Arg192Ser was nominally associated with T2D (OR 1.62, $P = 5.18 \times 10^{-4}$). In T2D patients, *PAX4* Arg192His was associated with earlier age at diagnosis, and *GLP1R* Arg131Gln was associated with decreased risk of cardiovascular disease. In control subjects without diabetes, the *PAX4* Arg192His was associated with higher fasting glucose and *GLP1R* Arg131Gln was associated with lower fasting glucose and HbA$_{1c}$ level. Gene-based analysis revealed that *SLC30A8* was most significantly associated with decreased risk of T2D ($P = 1.0 \times 10^{-4}$). In summary, we have identified nonsynonymous variants associated with risk of T2D and related phenotypes in Koreans.

It has been recognized that there are differences in clinical characteristics of type 2 diabetes (T2D) between East Asians and Europeans (1). East Asian patients with T2D

[1]Department of Internal Medicine, Seoul National University Hospital, Seoul, Republic of Korea

[2]Department of Biochemistry and Molecular Biology, Seoul National University College of Medicine, Seoul, Republic of Korea

[3]Research Institute of Basic Sciences, Seoul National University, Seoul, Republic of Korea

[4]Department of Internal Medicine, Seoul National University Boramae Medical Center, Seoul, Republic of Korea

[5]Department of Internal Medicine, Healthcare System Gangnam Center, Seoul National University Hospital, Seoul, Republic of Korea

[6]Department of Family Medicine, Seoul National University Hospital, Seoul, Republic of Korea

[7]Department of Internal Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea

[8]Department of Internal Medicine, Seoul National University Bundang Hospital, Seongnam, Republic of Korea

[9]Division of Genome Research, Center for Genome Science, Korea National Institute of Health, Cheongju, Republic of Korea

[10]Department of Biomedical Science, Hallym University, Chuncheon, Republic of Korea

[11]Severance Biomedical Science Institute and Department of Internal Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea

[12]Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, MI

[13]Department of Statistics, Seoul National University, Seoul, Republic of Korea

[14]Department of Preventive Medicine and Public Health, Ajou University School of Medicine, Suwon, Republic of Korea

[15]Statistics and Data Center, Research Institute for Future Medicine, Samsung Medical Center, Seoul, Republic of Korea

[16]Department of Digital Health, Samsung Advanced Institute for Health Sciences and Technology, Sungkyunkwan University, Seoul, Republic of Korea

[17]Department of Molecular Medicine and Biopharmaceutical Sciences, Graduate School of Convergence Science and Technology, Seoul National University, Seoul, Republic of Korea

Corresponding author: Kyong Soo Park, kspark@snu.ac.kr.

tend to have relatively lower BMI, earlier age at diagnosis, and decreased insulin secretion (2,3). In Europeans, it is widely accepted that obesity and the resulting increased insulin resistance play a crucial role in the pathophysiology (4). In contrast, East Asians have β-cell dysfunction that precedes the development of diabetes and their β-cell compensation is limited (3). It has been suggested that these differences could be the result of both genetic and environmental factors specific to each population.

Genetic variations may influence the ethnicity-specific clinical characteristics of T2D in several ways. There are common variants that exert heterogeneous allelic effects according to ethnic group (5). In addition, the haplotype structure of certain risk loci and variant allele frequency are different among populations (6,7). Although previous common variant genome-wide association studies were successful in identifying more than 100 genetic risk loci of T2D, there were limitations to investigating ethnicity-specific variants. Advances in genome sequencing technologies have enabled us to investigate rare coding variants that are specific to a population. A recent large-scale collaborative study identified a number of rare coding variants of T2D (8). It was shown that rare variants had a limited role in the overall predisposition to T2D. Still, a significant proportion of rare coding variants are restricted to closely related populations (9), and there are reports that isolated populations could be useful for identifying rare functional variants of common complex disorders (10). It would be of interest to investigate ethnicity-specific coding variants that predispose to T2D and affect related clinical phenotypes in East Asians.

In this study, we investigated coding variants that are associated with risk of T2D in 3,943 Korean case and control subjects using whole-exome sequencing and exome array genotyping. The findings were validated by in silico replication in an additional 13,122 case and control subjects. We also investigated how these variants are related to T2D-related phenotypes in Koreans.

## RESEARCH DESIGN AND METHODS

### Study Participants

This was a three-staged case-control study, with stage 1 whole-exome sequencing involving 917 participants (619 T2D case subjects and 298 control subjects without diabetes), stage 2 exome array genotyping in 3,026 participants (2,013 case and 1,013 control subjects), and stage 3 in silico replication for seven nonsynonymous variants in 13,122 participants (5,218 case and 7,904 control subjects) from the Korean population. The overall study design, including the number of participants in each stage and variants analyzed, is shown in Fig. 1. The clinical characteristics of the stage 1 and 2 study participants are shown in Table 1, and those of stage 3 participants are shown in Supplementary Table 1. T2D patients of stage 1 and 2 analyses were enrolled from Seoul National University Hospital between January 2001 and September

2014. T2D was diagnosed according to the recent American Diabetes Association guidelines (11). T2D case subjects were preferentially enrolled if there were any first-degree relatives with T2D. Control subjects of stage 1 and 2 were enrolled if they were aged ≥60 years, had a fasting glucose <100 mg/dL, had an $HbA_{1c}$ <6.0%, and had no first-degree relatives with T2D. Inclusion criteria for each study comprising the stage 3 replication are shown in Supplementary Table 1. The institutional review board of the Biomedical Research Institute at Seoul National University Hospital approved the study protocol (1205–130–411), and written informed consent was obtained from each participant. All clinical investigations were conducted according to the Declaration of Helsinki.

### Stage 1: Whole-Exome Sequencing

A total of 917 samples were used for stage 1 whole-exome sequencing. Whole-exome sequencing was performed as previously described (12). In brief, genomic DNA was used for exome capture using Agilent SureSelect v4+UTR, and massive parallel sequencing was performed using Illumina HiSeq 2000 sequencing system by Macrogen, Inc. (Seoul, Republic of Korea). The minimum average depth of coverage for target region was 80×. Sequence reads were aligned to the human reference genome (GRCh37) and processed using BWA (13), Picard (http://broadinstitute .github.io/picard/), and Genome Analysis Toolkit software (14). Variants were called using multisample HaplotypeCaller of Genome Analysis Toolkit (14). The summary statistics of sequence read output are displayed in Supplementary Table 2. The number of total variants and per individual average counts of the variants in the sequenced subjects are displayed in Supplementary Tables 3 and 4. The variant quality-control measures of transition-to-transversion ratio and heterozygous-to-homozygous ratio are summarized in Supplementary Table 5. Among the stage 1 samples, DNA was available from 858 participants for further common variant analysis using either Genome-Wide Human SNP Array 5.0 (Affymetrix, Santa Clara, CA) ($N$ = 504) or Axiom Biobank Plus Genotyping Array (Affymetrix) ($N$ = 354). Data from the two arrays were merged for further analysis after excluding variants with a significant difference in missingness ($P < 1.0 \times 10^{-3}$).

### Stage 2: Exome Array Genotyping

Stage 2 exome array genotyping was performed using a total of 3,026 independent samples. Exome array genotyping was performed with the Affymetrix Axiom Biobank Plus Genotyping Array by DNA Link, Inc. (Seoul, Republic of Korea). We customized the array by incorporating 144,339 variants that were either nonsynonymous or had a putative association with T2D or related phenotypes with a significance threshold of $P < 0.10$ in the stage 1 analysis. Samples with missing genotype rate >5% or sex inconsistency were removed. Variants that showed deviation from Hardy-Weinberg equilibrium ($P < 1.0 \times 10^{-6}$), had a genotype call rate <95%, or were monomorphic were excluded. A total of 369,694 variants were
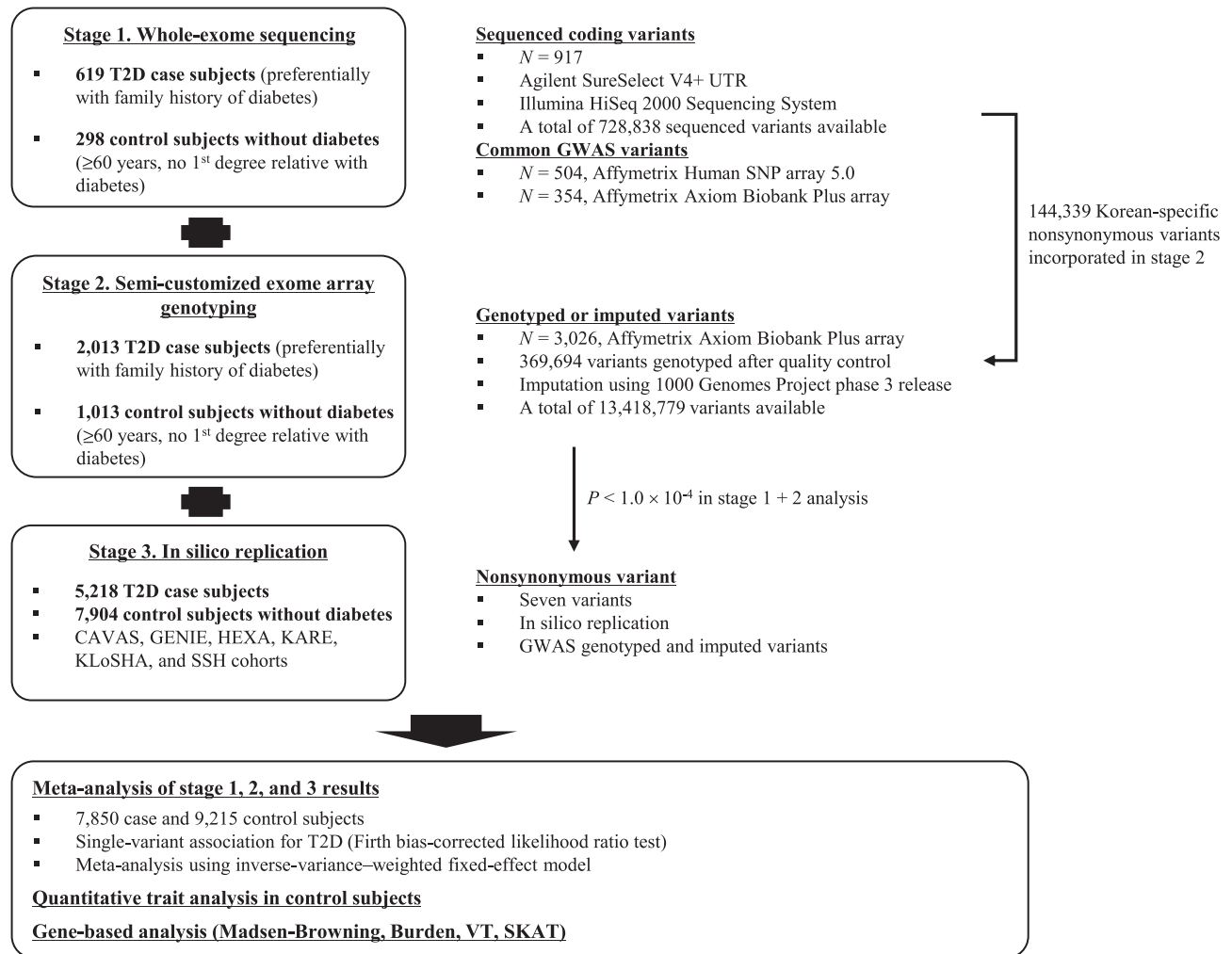
**Figure 1**—The overall study design, including the number of participants in each stage and variants analyzed. This was a three-staged case-control study with stage 1 whole-exome sequencing involving 917 participants, stage 2 exome array genotyping in 3,026 participants, and stage 3 in silico replication for seven nonsynonymous variants in 13,122 participants. GWAS, genome-wide association study; SKAT, sequence kernel association test; VT, variable threshold method.

available for analysis. The genotype concordance rate in the 13 replicated samples was 99.9%. After quality control, 3,026 samples were available for stage 2 analysis. Genotype imputation was performed using the 1000 Genomes Project phase 3 release reference panel (9). Imputation was conducted with IMPUTE2 software according to the best practice guideline (https://mathgen.stats.ox.ac.uk/impute) (15). Postimputation quality control using INFO score $\geq$0.4 was applied, and a total of 13,418,779 variants were available for analysis.

**Stage 3: In Silico Replication**

Seven nonsynonymous variants that showed association with predefined threshold of $P < 1.0 \times 10^{-4}$ in stage 1 and 2 meta-analyses were evaluated by data look-up in stage 3 in silico replication. Stage 3 in silico replication was conducted using six cohorts encompassing 13,122 participants. The six studies were Cardiovascular Disease Association Study (CAVAS), Seoul National University Hospital

Gene-ENvironment of Interaction and Phenotype (GENIE) Study, Health Examinees (HEXA) shared control study, Korea Association REsource (KARE) project, Korean Longitudinal Study on Health and Aging (KLoSHA), and Samsung Hospital (SSH) cohort. The detailed information of each study including genotyping platform is provided in Supplementary Table 1.

**Single-Variant Association Test**

The statistical power of this study was estimated for single variants with a genetic power calculator (http://zzz.bwh.harvard.edu/gpc), assuming T2D prevalence of 10%, risk allele frequency 5%, and genome-wide significance of $\alpha < 5.0 \times 10^{-8}$ in a 2 degrees of freedom additive genetic model (16). For variants with odds ratios (ORs) of 1.3 and 1.4, this study had a power of 79.3% and 99.7%, respectively. Single-variant association testing for T2D was performed with EPACTS software (http://genome.sph.umich.edu/wiki/EPACTS). Throughout the article, alternative

**Table 1—Clinical characteristics of study participants of stage 1 and 2 analyses (N = 3,943)**

| | Stage 1: whole-exome sequencing | | | Stage 2: exome array genotyping | | |
|---|---|---|---|---|---|---|
| | T2D subjects | Control subjects | *P* | T2D subjects | Control subjects | *P* |
| *N* | 619 | 298 | | 2,013 | 1,013 | |
| Male/female, *n* | 287/332 | 133/165 | 0.672 | 933/1,080 | 515/498 | 0.021 |
| Age (years) | 56.4 ± 9.1 | 66.9 ± 7.1 | <0.001 | 58.4 ± 13.5 | 65.5 ± 8.3 | <0.001 |
| Onset age (years) | 48.5 ± 10.1 | NA | NA | 47.0 ± 12.0 | NA | NA |
| BMI (kg/m$^2$) | 24.4 ± 2.8 | 23.7 ± 3.2 | <0.001 | 24.3 ± 3.2 | 23.8 ± 3.0 | <0.001 |
| Systolic blood pressure (mmHg) | 127 ± 19 | 130 ± 20 | 0.011 | 129 ± 18 | 125 ± 16 | <0.001 |
| Diastolic blood pressure (mmHg) | 77 ± 10 | 81 ± 12 | <0.001 | 78 ± 11 | 78 ± 10 | 0.608 |
| HbA$_{1c}$ (%) | 8.1 ± 1.8 | 5.2 ± 0.4 | <0.001 | 7.7 ± 1.6 | 5.5 ± 0.3 | <0.001 |
| HbA$_{1c}$ (mmol/mol) | 65 ± 19.7 | 33 ± 4.4 | <0.001 | 61 ± 17.5 | 37 ± 3.3 | <0.001 |
| Fasting glucose (mg/dL) | 152 ± 49 | 91 ± 6 | <0.001 | 147 ± 52 | 90 ± 8 | <0.001 |
| Total cholesterol (mg/dL) | 196 ± 41 | 199 ± 33 | 0.218 | 178 ± 40 | 194 ± 35 | <0.001 |
| Triglyceride (mg/dL) | 163 ± 128 | 129 ± 74 | <0.001 | 149 ± 94 | 118 ± 65 | <0.001 |
| HDL cholesterol (mg/dL) | 48 ± 12 | 46 ± 16 | 0.123 | 48 ± 14 | 53 ± 14 | <0.001 |
| LDL cholesterol (mg/dL) | 118 ± 33 | 128 ± 30 | <0.001 | 104 ± 34 | 120 ± 32 | <0.001 |
| Insulin treatment, % | 26.8 | NA | NA | 23.8 | NA | NA |
| Lipid medication, % | 21.0 | 0.0 | <0.001 | 27.5 | 2.7 | <0.001 |
| Hypertension medication, % | 36.7 | 7.0 | <0.001 | 32.0 | 10.0 | <0.001 |

Data are expressed as the mean ± SD, unless stated otherwise. NA, not applicable.

alleles are reported as effect alleles. Hardy-Weinberg equilibrium was tested as implemented in EPACTS, and markers with significant deviation ($P < 1.0 \times 10^{-6}$) were excluded from the analysis. The Firth bias–corrected logistic likelihood ratio model (17) was used with covariates of age, sex, and four principal components to adjust for population stratification. Principal components were estimated with linkage disequilibrium–pruned ($r^2 < 0.2$) exome array variants with a minor allele frequency (MAF) of 5%. The stage 1, 2, and 3 results were meta-analyzed using METAL software with an inverse-variance–weighted fixed-effects model (18). The threshold for genome-wide significant association was set as $P < 5.0 \times 10^{-8}$. Glycemic traits were analyzed in participants without diabetes of the KARE cohort ($N = 7,516$). The associations between genetic variants and fasting glucose, 1-h and 2-h glucose of 75-g oral glucose tolerance test, fasting insulin, and HbA$_{1c}$ were analyzed using linear regression adjusted for age and sex, assuming additive genetic model. Continuous variables were inverse normal transformed before analysis. Further genetic association with risk of cardiovascular disease (CVD) in T2D case subjects was performed in a subset of Seoul National University Hospital participants ($N = 1,496$). Information on previous history of CVD was collected by patient report and review of medical records. CVD events included stable or unstable angina, myocardial infarction, history of percutaneous coronary intervention, coronary artery bypass surgery, and ischemic or hemorrhage stroke. Participants with any of the above events were categorized as CVD case subjects and those without were

classified as non-CVD control subjects. The clinical characteristics of participants analyzed for CVD events are shown in Supplementary Table 6. Genetic association testing was performed using logistic regression with adjustment for age and sex.

**Gene-Based Association Test**
A gene-based analysis was performed with RAREMETAL software in 3,943 participants of stage 1 and 2 results (19). Variants were annotated using ANNOVAR software (20). Variants predicted to result in protein truncation (stop gain or loss, frameshift, essential splice site) or nonsynonymous variants with a MAF <0.01 were used for gene-based tests. Four gene-based burden tests, which included the Madsen-Browning method (21), the combined multivariate and collapsing method (22), the variable threshold method (23), and the sequence kernel association test (24), were performed using the whole-exome sequence data from the stage 1 analysis and the imputed genotype data from the stage 2 analysis. A meta-analysis of the stage 1 and 2 results was performed as implemented in the RAREMETAL software.

A gene-based test by aggregating coding variants was performed for 28 genes that are implicated in monogenic diabetes, such as maturity-onset diabetes of the young and neonatal diabetes (Supplementary Table 7) (8). Variants were grouped into three masks: mask 1, variants annotated as high-confidence disease-causing mutation in the Human Gene Mutation Database (HGMD) or protein-truncating variants; mask 2, mask 1 plus variants annotated to be deleterious in all five prediction models of PolyPhen2-HumDiv,

**Table 2—Association of nonsynonymous variants with risk of T2D (N = 17,065)**

| Chr | Position (hg19) | Alleles (Ref, Alt) | Gene | HGVS (rsID) | Korean AF | European AF | Study | OR (95% CI) | P | AF case subjects | AF control subjects |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 7 | 127253550 | C, T | PAX4 | p.Arg192His (rs2233580) | 0.086 | 0.000 | Stage 1. WES | 2.62 (1.64–4.18) | $1.73 \times 10^{-5}$ | 0.122 | 0.057 |
|  |  |  |  |  |  |  | Stage 2. Exome chip | 1.67 (1.33–2.08) | $3.16 \times 10^{-6}$ | 0.094 | 0.059 |
|  |  |  |  |  |  |  | Stage 3. Replication | 1.40 (1.26–1.56) | $9.22 \times 10^{-10}$ | — | — |
|  |  |  |  |  |  |  | Meta-analysis | **1.48 (1.35–1.63)** | $\mathbf{4.47 \times 10^{-16}}$ | — | — |
| 6 | 39033595 | G, A | GLP1R | p.Arg131Gln (rs3765467) | 0.211 | 0.001 | Stage 1. WES | 0.74 (0.55–0.98) | 0.036 | 0.196 | 0.258 |
|  |  |  |  |  |  |  | Stage 2. Exome chip | 0.78 (0.69–0.90) | $3.60 \times 10^{-4}$ | 0.198 | 0.234 |
|  |  |  |  |  |  |  | Stage 3. Replication | 0.87 (0.81–0.93) | $5.77 \times 10^{-5}$ | — | — |
|  |  |  |  |  |  |  | Meta-analysis | **0.84 (0.80–0.90)** | $\mathbf{3.55 \times 10^{-8}}$ | — | — |
| 4 | 1349029 | G, A | UVSSA | p.Arg391His (rs2276904) | 0.376 | 0.031 | Stage 1. WES | 0.91 (0.72–1.15) | 0.441 | 0.363 | 0.393 |
|  |  |  |  |  |  |  | Stage 2. Exome chip | 0.77 (0.68–0.86) | $5.52 \times 10^{-6}$ | 0.357 | 0.417 |
|  |  |  |  |  |  |  | Stage 3. Replication | 0.94 (0.89–1.00) | 0.038 | — | — |
|  |  |  |  |  |  |  | Meta-analysis | 0.91 (0.86–0.95) | $8.36 \times 10^{-5}$ | — | — |
| 10 | 64974537 | A, T | JMJD1C | p.Ser464Thr (rs10761725) | 0.430 | 0.778 | Stage 1. WES | 1.34 (1.05–1.71) | 0.016 | 0.460 | 0.379 |
|  |  |  |  |  |  |  | Stage 2. Exome chip | 1.21 (1.08–1.36) | 0.001 | 0.443 | 0.401 |
|  |  |  |  |  |  |  | Stage 3. Replication | 1.07 (1.01–1.13) | 0.014 | — | — |
|  |  |  |  |  |  |  | Meta-analysis | 1.11 (1.05–1.16) | $6.14 \times 10^{-5}$ | — | — |
| 19 | 18123738 | T, C | ARRDC2 | p.Leu391Pro (rs7259041) | 0.251 | 0.254 | Stage 1. WES | 0.90 (0.69–1.17) | 0.424 | 0.239 | 0.268 |
|  |  |  |  |  |  |  | Stage 2. Exome chip | 0.77 (0.68–0.87) | $4.97 \times 10^{-5}$ | 0.236 | 0.280 |
|  |  |  |  |  |  |  | Stage 3. Replication | 0.95 (0.89–1.02) | 0.149 | — | — |
|  |  |  |  |  |  |  | Meta-analysis | 0.91 (0.85–0.96) | 0.001 | — | — |
| 11 | 2869129 | G, A | KCNQ1 | p.Gly643Ser (rs1800172) | 0.053 | 0.000 | Stage 1. WES | 0.73 (0.44–1.21) | 0.217 | 0.048 | 0.055 |
|  |  |  |  |  |  |  | Stage 2. Exome chip | 0.63 (0.49–0.80) | $1.58 \times 10^{-4}$ | 0.046 | 0.068 |
|  |  |  |  |  |  |  | Stage 3. Replication | 0.99 (0.86–1.13) | 0.835 | — | — |
|  |  |  |  |  |  |  | Meta-analysis | 0.84 (0.74–0.95) | 0.006 | — | — |
| 12 | 71533622 | C, T | TSPAN8 | p.Gly44Ser (rs79443892) | 0.157 | 0.018 | Stage 1. WES | 0.68 (0.51–0.92) | 0.011 | 0.137 | 0.188 |
|  |  |  |  |  |  |  | Stage 2. Exome chip | 0.78 (0.67–0.91) | 0.001 | 0.146 | 0.183 |
|  |  |  |  |  |  |  | Stage 3. Replication | 1.07 (1.00–1.15) | 0.062 | — | — |
|  |  |  |  |  |  |  | Meta-analysis | 0.98 (0.92–1.05) | 0.582 | — | — |

OR and P values are from T2D association testing results for alternative alleles adjusted for age, sex, and principle components. Alternative alleles are reported as effect alleles. Stage 1, 2, and 3 associations were analyzed with the Firth bias–corrected likelihood ratio test. Meta-analysis was performed using METAL with the inverse-variance–weighted method under a fixed-effects model. Alleles are aligned to the forward strand of the Human Genome Version 19 (hg19). Alt, alternative allele; AF, allele frequency; Chr, chromosome; HGVS, Human Genome Variation Society nomenclature; Ref, reference allele; rsID, dbSNP reference ID; WES, whole-exome sequencing. The genome-wide significance threshold was set to $P < 5.0 \times 10^{-8}$, shown in boldface type.

PolyPhen2-HumVar, LRT, MutationTaster, and SIFT (NS$_{strict}$) (25); and mask 3, mask 1 plus variants annotated to be deleterious in any of the five prediction models of PolyPhen2-HumDiv, PolyPhen2-HumVar, LRT, MutationTaster, and SIFT (NS$_{broad}$). Each of the three masks was also subdivided according to MAF cutoff of <0.001, <0.005, and <0.010, respectively. Logistic regression tests using a collapsing burden method by encoding individuals as 0 (if carrying no variant) or 1 (if carrying at least one variant) were performed for each combination of variant mask and MAF bin.

## RESULTS

### Overall Single-Variant Association

In the stage 1 whole-exome sequencing, a total of 728,838 variants (673,667 single nucleotide variants and 55,171 insertion-deletion variants) were investigated in 917 T2D case and control subjects. A total of 144,339 variants that were either nonsynonymous or had a putative association with T2D or related phenotypes in stage 1 analysis were incorporated in the exome array. In the stage 2 exome array genotyping, 369,694 variants were successfully genotyped in 3,026 T2D case and control subjects. The number of variants that were available after imputation in stage 2 was 13,418,779. An inverse-variance fixed-effects meta-analysis of stage 1 and 2 results was performed. The quantile-quantile plot and Manhattan plot are displayed in Supplementary Fig. 1. Four common noncoding variants in previously established loci, namely CDKAL1, CDKN2A/2B, KCNQ1, and MAEA, were confirmed to be associated with T2D at genome-wide significance (Supplementary Table 8).

Among the nonsynonymous variants, seven had an association with T2D with the predefined arbitrary significance threshold of $P < 1.0 \times 10^{-4}$ after stage 1 and 2 meta-analysis (Table 2). These variants were further investigated by in silico data look-up in an additional 13,122 participants in the stage 3 replication analysis (Supplementary Table 9). Four variants reached nominal significance of $P < 0.05$ in stage 3 analysis as shown in Table 2. Among these, two variants (PAX4 Arg192His [rs2233580] and GLP1R Arg131Gln [rs3765467]) reached genome-wide significance of $P < 5.0 \times 10^{-8}$. These two variants were common nonsynonymous variants with a MAF of 0.086 and 0.211, respectively, in Koreans. According to the 1000 Genomes Project database, they were also common in Han Chinese in Beijing (MAF of 0.093 and 0.218, respectively) and Japanese in Tokyo (MAF of 0.089 and 0.197, respectively) (9). However, these variants were absent or rare in Europeans (MAF <0.001). We further investigated the genetic association of 21 coding variants that were previously confirmed to be associated with T2D in Europeans (8) using our stage 1 and 2 participants (Supplementary Table 10). Although most of the associated variants showed directional consistency, we found that eight variants were monomorphic in our population.

**Table 3—Association of nonsynonymous variants with glycemic traits in control subjects without diabetes (N = 7,516)**

| Gene | HGVS (rsID) | Fasting glucose β (95% CI) | P | 1-h glucose β (95% CI) | P | 2-h glucose β (95% CI) | P | Fasting insulin β (95% CI) | P | HbA1c β (95% CI) | P |
|---|---|---|---|---|---|---|---|---|---|---|---|
| PAX4 | p.Arg192His (rs2233580) | 0.103 (0.036–0.170) | **2.68 × 10⁻³** | 0.071 (0.004–0.138) | **0.038** | −0.038 (−0.105, 0.029) | 0.263 | 0.039 (−0.028, 0.106) | 0.258 | 0.029 (−0.038, 0.096) | 0.402 |
| GLP1R | p.Arg131Gln (rs3765467) | −0.075 (−0.114, −0.035) | **2.00 × 10⁻⁴** | −0.048 (−0.088, −0.009) | **0.016** | −0.052 (−0.091, −0.012) | **0.01** | −0.031 (−0.071, 0.008) | 0.12 | −0.053 (−0.093, −0.014) | **8.52 × 10⁻³** |
| UVSSA | p.Arg391His (rs2276904) | −0.001 (−0.033, 0.032) | 0.956 | −0.024 (−0.056, 0.009) | 0.154 | 0.009 (−0.023, 0.042) | 0.575 | −0.018 (−0.051, 0.014) | 0.274 | −0.016 (−0.048, 0.017) | 0.349 |
| JMJD1 | p.Ser464Thr (rs10761725) | 0.014 (−0.019, 0.046) | 0.415 | 0.059 (0.026–0.091) | **3.90 × 10⁻⁴** | 0.047 (0.014–0.079) | **4.62 × 10⁻³** | −0.02 (−0.053, 0.012) | 0.22 | −0.005 (−0.037, 0.028) | 0.783 |

β and P values are from linear regression analysis for alternative alleles adjusted for age and sex, using additive genetic model. Alternative alleles are reported as effect alleles. Variables were inverse normal transformed before analysis. HGVS, Human Genome Variation Society nomenclature; rsID, dbSNP reference ID. Boldface type denotes P < 0.05.

**Table 4—Amino acid variation in the *PAX4* 192 codon and risk of T2D in stage 1 and 2 analyses**

| *PAX4* 192 codon | Mutant codons, *N* | T2D subjects, *N* (%) | Control subjects, *N* (%) | OR | 95% CI | *P* |
|---|---|---|---|---|---|---|
| Arg/Arg | 0 | 1,817 (72.3) | 1,054 (83.1) | 1.00 | — | — |
| Arg/His, Arg/Ser | 1 | 635 (25.3) | 203 (16.0) | 1.78 | 1.49–2.15 | $6.79 \times 10^{-10}$ |
| His/Ser, His/His, Ser/Ser | 2 | 61 (2.4) | 11 (0.9) | 3.23 | 1.64–6.35 | $6.95 \times 10^{-4}$ |

OR and *P* values are from logistic regression analysis adjusting for age and sex. Participants with Arg/Arg were used as the reference.

### *PAX4* Amino Acid Variation

Among the nonsynonymous variants, the *PAX4* Arg192His variant (chr7:127253550, rs2233580, effect allele T) was most significantly associated with increased risk of T2D in Koreans, with an OR of 1.48 (95% CI 1.35–1.63, $P = 4.47 \times 10^{-16}$). This Arg192His variant (T allele) was associated with increased fasting glucose and 1-h glucose in 7,516 control subjects without diabetes, as shown in Table 3 ($P < 0.05$). T2D participants harboring this variant had a significantly younger age at diagnosis and lower C-peptide levels ($P < 0.05$) (Supplementary Table 11). We found an independent low-frequency nonsynonymous variant (chr7:127253551, rs3824004, effect allele T) that resulted in a different amino acid change in the *PAX4* 192 codon (Arg192Ser) in meta-analysis of stage 1 and 2 results. However, this variant was not well imputed in the stage 3 analysis. The *PAX4* Arg192Ser variant was also associated with risk of T2D, with an OR of 1.62 (95% CI 1.41–1.86, $P = 5.18 \times 10^{-4}$). The two variants were not in linkage disequilibrium ($r^2 = 0.002$). The *PAX4* 192 amino acid was determined by the combination of rs2233580 and rs3824004: Arg/Arg, Arg/His, Arg/Ser, His/Ser, His/His, and Ser/Ser (Table 2). Having one of either amino acid change (Arg/His or Arg/Ser) increased the risk of T2D compared with wild-type Arg/Arg, with an OR of 1.78 (95% CI 1.49–2.15, $P = 6.79 \times 10^{-10}$) in the combined stage 1 and 2 analyses. Having two of the amino acid changes (His/Ser, His/His, or Ser/Ser) increased the risk of T2D, with an OR of 3.23 (95% CI 1.64–6.35, $P = 6.95 \times 10^{-4}$) compared with wild type (Table 4).

### *GLP1R* Amino Acid Variation

A nonsynonymous variant in *GLP1R* Arg131Gln (chr6: 39033595, rs3765467, effect allele A) was significantly associated with decreased risk of T2D (OR 0.84, 95% CI 0.80–0.90, $P = 3.55 \times 10^{-8}$). The Arg131Gln variant (A allele) was also associated with decreased fasting glucose, 1-h and 2-h glucose, and decreased HbA$_{1c}$ in control subjects without diabetes ($P < 0.05$), as shown in Table 3. We further investigated whether this variant was associated with risk of CVD. In a subset of 1,496 T2D case subjects among stage 1 and 2 participants, this variant was nominally associated with decreased risk of CVD (OR 0.77, 95% CI 0.59–0.99, $P = 0.041$) (Table 5). Finally, gene-based association test using rare nonsynonymous variants showed that *GLP1R* was nominally associated with decreased risk of T2D according to Madsen-Browning weighted sum statistics method (effect = −0.0013, $P = 0.039$).

### Gene-Based Associations

We investigated gene-based associations with T2D using four statistical methods implemented in RAREMETAL (Table 6). None of the genes were associated with T2D at exome-wide significance threshold ($P < 2.5 \times 10^{-6}$). However, *SLC30A8* was most significantly associated with decreased risk of T2D according to Madsen-Browning weighted sum statistics method ($P = 1.00 \times 10^{-4}$). There were five rare nonsynonymous variants in this gene that were preferentially observed in control subjects without diabetes. Other genes that had suggestive association with the predefined arbitrary threshold of $P < 1.0 \times 10^{-3}$ included *DOCK9*, *PPAPDC1B*, *TMEM62*, *NUCB1*, *IAH1*, and *RAPGEF4* (*EPAC2*).

We further investigated the enrichment of rare deleterious variants in genes implicated in monogenic forms of diabetes. A total of 28 genes known to cause monogenic diabetes were included for analysis (8) (Supplementary Table 7). Logistic regression analysis using a collapsing burden method revealed that the effect size increased as more stringent criteria for functionality were used and as the MAF cutoff was decreased (Supplementary Fig. 2). For example, when variants were limited to HGMD disease-causing mutations or protein-truncating variants with MAF <0.001, they were significantly associated with T2D, with an OR of 2.81 (95% CI 1.08–7.29, $P = 0.042$).

### DISCUSSION

Using combined whole-exome sequencing, exome array genotyping, and in silico replication in a total of 17,065 Korean T2D case and control subjects, we confirmed that two nonsynonymous variants, *PAX4* Arg192His and *GLP1R* Arg131Gln, were significantly associated with T2D in the Korean population. We also found that *PAX4* Arg192Ser, another amino acid variant of *PAX4*, was associated with the risk of T2D. The combination of the two *PAX4* 192 codon variants was a strong genetic risk factor for T2D in Koreans. It was interesting to identify a nonsynonymous *GLP1R* Arg131Gln variant to be associated with decreased risk of T2D. This variant was also associated with decreased glucose concentration and protection from CVD. Gene-based testing also confirmed previous association results that rare nonsynonymous variants in *SLC30A8* were associated with protection from T2D. To the best of our knowledge, this is the first study in East Asians that used whole-exome sequencing, customized exome array genotyping, and in

**Table 5—Association of nonsynonymous variants with CVD (N = 1,496)**

| Gene | HGVS (rsID) | Effect allele | CVD case subjects, n | CVD control subjects, n | CVD case subjects, AF | CVD control subjects, AF | OR (95% CI) | P |
|---|---|---|---|---|---|---|---|---|
| PAX4 | p.Arg192His (rs2233580) | T | 239 | 1,257 | 0.125 | 0.096 | 1.30 (0.95–1.78) | 0.100 |
| GLP1R | p.Arg131Gln (rs3765467) | A | 239 | 1,257 | 0.159 | 0.199 | **0.77 (0.59–0.99)** | **0.041** |
| UVSSA | p.Arg39His (rs2276904) | A | 239 | 1,257 | 0.370 | 0.378 | 0.97 (0.79–1.18) | 0.748 |
| JMJD1 | p.Ser464Thr (rs10761725) | T | 239 | 1,257 | 0.417 | 0.445 | 0.83 (0.68–1.02) | 0.077 |

OR and P values are from CVD association testing results for alternative alleles adjusted for age and sex. A subset of T2D patients (N = 1,496) whose CVD event status was available from stage 1 and 2 analyses was investigated with logistic regression. Alternative alleles are reported as effect alleles. CVD events included stable or unstable angina, myocardial infarction, history of percutaneous coronary intervention, coronary artery bypass surgery, or ischemic or hemorrhagic stroke. AF, allele frequency; HGVS, Human Genome Variation Society nomenclature; rsID, dbSNP reference ID. Boldface type denotes P < 0.05.

silico replication to investigate the association of nonsynonymous variants with T2D.

One of the major aims of this study was to identify variants that were specific to Koreans and investigate their association with T2D. The sequencing was performed with high quality, with average depth of coverage >80× for each sample. Using whole-exome sequencing, we identified 232,355 novel single nucleotide variations and 26,108 novel insertion-deletions that were not captured in the database of Short Genetic Variations (dbSNP147) (Supplementary Table 3). All seven nonsynonymous variants listed in Table 2 were common, with a MAF ≥0.05 in East Asians. However, five of these were rare or even absent in Europeans. For example, the frequency of the PAX4 Arg192His variant was 0.086 in Koreans but was absent in Europeans. The variant frequency of GLP1R Arg131Gln was 0.211 in Koreans compared with 0.001 in Europeans. In addition, among the previously confirmed 21 coding variants that reached genome-wide significance in Europeans, 8 were monomorphic in East Asians. This suggests that the MAF of nonsynonymous variants is quite different among different populations and that associations between nonsynonymous variants and T2D could be population specific. These findings support previous observations of genetic heterogeneity between different ethnic groups, especially for coding variants. Therefore, it would be important to investigate ethnicity-specific genetic variations in order to translate genetic information into clinical practice.

The association between PAX4 Arg192His variants and T2D was previously reported (8). We expanded upon this finding by identifying that amino acid variations of the 192 codon determine the risk of T2D and the clinical characteristics of T2D in Koreans. Having one copy of the Arg192His variant was associated with a 1.4-year earlier onset of diabetes. Having two copies of the variant was associated with a 7.0-year earlier onset. This variant was associated with decreased C-peptide levels in T2D participants. One of the novel findings of this study was that another variant that resulted in PAX4 Arg192Ser was also associated with T2D, with at least nominal significance (P = $5.18 \times 10^{-4}$). This variant was in low frequency and specific to East Asians. The effect size of the two variants was similar, and the amino acid combinations were a strong risk factor for T2D. The PAX4 gene encodes paired box 4, which is a transcription factor that plays crucial role in pancreatic islet development and β-cell differentiation (26). It also represses glucagon gene expression (27). Rare mutations in PAX4 have been reported to cause monogenic diabetes in South Asians (28). Our study suggests that PAX4 Arg192His variant increases the predisposition to early-onset T2D rather than causing monogenic diabetes. However, the detailed mechanism of how this variant increases the risk of T2D is largely unknown and further functional investigations are warranted.

We have identified an ethnicity-specific common nonsynonymous variant in GLP1R that is associated with protection from T2D, lower glucose concentration, and

**Table 6—Genes associated with T2D by gene-based test in stage 1 and 2 analyses**

| Gene | Chr | Variants, n | Madsen–Browning | | Burden | | Variable threshold | | SKAT | Protein | Biological function (UniProtKB) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Effect | P | Effect | P | Effect | P | P | | |
| SLC30A8 | 8 | 5 | −0.02 | $1.00 \times 10^{-4}$ | −0.46 | 0.0016 | −0.90 | 0.0031 | 0.105 | Zinc transporter 8 | Facilitates the accumulation of zinc from the cytoplasm into intracellular vesicles as a zinc-efflux transporter. May be a major component that provides zinc for insulin maturation and/or storage processes in insulin-secreting pancreatic β-cells. |
| DOCK9 | 13 | 13 | −0.02 | $2.92 \times 10^{-4}$ | −0.40 | 0.0024 | −1.10 | 0.0019 | 0.211 | Dedicator of cytokinesis protein 9 | Guanine nucleotide-exchange factor that activates CDC42 by exchanging bound GDP for free GTP. Overexpression induces filopodia formation. |
| PPAPDC1B | 8 | 3 | −0.03 | $3.23 \times 10^{-4}$ | −0.70 | $9.88 \times 10^{-4}$ | −1.33 | $1.46 \times 10^{-4}$ | 0.016 | Phospholipid phosphatase 5 | Displays magnesium-independent phosphatide phosphatase activity in vitro. Catalyzes the conversion of phosphatidic acid to diacylglycerol. May be a metastatic suppressor for hepatocellular carcinoma. |
| TMEM62 | 15 | 3 | −0.03 | $4.52 \times 10^{-4}$ | −0.89 | $7.72 \times 10^{-4}$ | −0.89 | 0.0020 | 0.017 | Transmembrane protein 62 | Function not known. |
| NUCB1 | 19 | 6 | −0.02 | $6.43 \times 10^{-4}$ | −0.66 | 0.0023 | −1.31 | 0.0046 | 0.112 | Nucleobindin-1 | Major calcium-binding protein of the Golgi. May have a role in calcium homeostasis. |
| IAH1 | 2 | 6 | −0.02 | $7.31 \times 10^{-4}$ | −0.97 | 0.0010 | −1.31 | 0.0039 | 0.136 | Isoamyl acetate-hydrolyzing esterase 1 homolog | Probable lipase. |
| RAPGEF4 (EPAC2) | 2 | 14 | 0.01 | $7.38 \times 10^{-4}$ | 0.38 | 0.0021 | 0.58 | 0.0089 | 0.347 | Rap guanine nucleotide exchange factor 4 | Guanine nucleotide exchange factor for RAP1A, RAP1B, and RAP2A small GTPases. Activated by binding cAMP. Seems to not activate RAB3A. Involved in cAMP-dependent, PKA-independent exocytosis through interaction with RIMS2. |

Effect and P values are from gene-based T2D association test, adjusting for age, sex, and principle components for each of the four statistical methods as implemented in RAREMETAL software. Genes were listed if statistical significance met the predefined arbitrary threshold of $P < 1.0 \times 10^{-3}$ for the Madsen-Browning method. The exome-wide significance threshold was set to $P < 2.5 \times 10^{-6}$. Biological function of the protein was referred from UniProt Knowledgebase (UniProtKB) (http://www.uniprot.org/help/uniprotkb). Chr, chromosome; SKAT, sequence kernel association test.

decreased risk of CVD in Koreans. The *GLP1R* encodes GLP-1 receptor and mediates the incretin effect exerted by GLP-1 and its therapeutic analogs. Recent randomized clinical trials showed that GLP-1 analogs resulted in decreased adverse CVD outcomes of cardiovascular death, nonfatal myocardial infarction, and nonfatal stroke in T2D patients (29,30). The 131 codon is located at a region where it links extracellular domain and transmembrane domain of GLP-1 receptor (31). The detailed alteration in GLP-1 signaling ensued by this variant is not well understood. However, the variant is near the peptide binding site of GLP-1 receptor (31). The *GLP1R* Arg131Gln variant has been shown to be associated with enhanced insulin secretion after intravenous GLP-1 infusion in control subjects without diabetes (32). In addition, we have reported that T2D patients harboring this variant had greater $HbA_{1c}$ reduction after dipeptidyl peptidase 4 inhibitor treatment (33). Asians have a better response to dipeptidyl peptidase 4 inhibitors compared with Europeans (34). This could be attributable, at least in part, to the genetic effect of the Arg131Gln, which is common in East Asians. This variant is very rare in Europeans, with a MAF = 0.001. However, it has been reported that a low-frequency variant (Ala316Thr, rs10305492) in *GLP1R* in Europeans was associated with decreased fasting glucose, decreased risk of T2D, and CVD (35). These lines of evidence consistently support that a nonsynonymous variant in *GLP1R* is significantly associated with risk of T2D and CVD.

Gene-based testing could be used to identify potential drug targets for diabetes. We confirmed the previous finding that rare coding variants of *SLC30A8*, in aggregate, are associated with a lower risk of T2D (36). It is plausible that a novel therapeutic agent that impairs *SLC30A8* function might have a protective role in T2D. Another gene of potential interest was *RAPGEF4*, also known as *EPAC2*. In gene-based testing, rare functional variants of this gene were associated with an increased risk of T2D. *EPAC2* is involved in cAMP-mediated signal transduction in pancreatic β-cells and directly interacts with sulfonylureas (37). It is expected that an EPAC2 agonist could stimulate insulin secretion in a glucose-dependent manner (38).

There were several limitations to this study. First, the sample size was modest for a whole-exome sequencing study. We had limited statistical power to identify rare variants associated with T2D and most of the significant findings were in common variants. The results of our findings should be interpreted cautiously, and the possibility of both type 1 and type 2 error should be considered. Second, the stage 2 samples were genotyped for selected coding variants, and stage 3 in silico replication was done for only seven variants. It could be possible that a number of variants might not have passed the criteria to be carried on for stage 2 and 3 investigations. In addition, low-frequency or rare variants might not have been genotyped correctly in the genotyping array. Third, there was an imbalance in the number of case and control subjects for stage 1 and 2 analyses. We sequenced and genotyped more

case than control subjects. This might have also influenced the statistical power.

In conclusion, using whole-exome sequencing, exome array genotyping, and in silico replication, we have identified that amino acid variations in *PAX4* and *GLP1R* are important genetic risk factors for T2D in Koreans. The *PAX4* Arg192His variant was associated with decreased C-peptide and younger age at diagnosis in T2D patients. A common nonsynonymous *GLP1R* Arg131Gln variant was associated with decreased risk of T2D, lower glucose concentration, and protection from CVD. This ethnicity-specific genomic information could be used to accelerate the realization of precision medicine in T2D.

### References

1.  Min HK. Non-insulin-dependent diabetes mellitus (NIDDM) in Korea. Diabet Med 1996;13(Suppl. 6):S13–S15
2.  Yoon KH, Lee JH, Kim JW, et al. Epidemic obesity and type 2 diabetes in Asia. Lancet 2006;368:1681–1688
3.  Ohn JH, Kwak SH, Cho YM, et al. 10-year trajectory of β-cell function and insulin sensitivity in the development of type 2 diabetes: a community-based prospective cohort study. Lancet Diabetes Endocrinol 2016;4:27–34
4.  Tabák AG, Jokela M, Akbaraly TN, Brunner EJ, Kivimäki M, Witte DR. Trajectories of glycaemia, insulin sensitivity, and insulin secretion before diagnosis of type 2 diabetes: an analysis from the Whitehall II study. Lancet 2009;373:2215–2221
5.  Mahajan A, Go MJ, Zhang W, et al.; DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium; Asian Genetic Epidemiology Network Type 2 Diabetes (AGEN-T2D) Consortium; South Asian Type 2 Diabetes (SAT2D) Consortium; Mexican American Type 2 Diabetes (MAT2D) Consortium; Type 2 Diabetes Genetic Exploration by Nex-generation sequencing in muylti-Ethnic Samples (T2D-GENES) Consortium. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nat Genet 2014;46:234–244
6.  Ng MC, Park KS, Oh B, 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 2008;57:2226–2233
7.  Sim X, Ong RT, Suo C, et al. Transferability of type 2 diabetes implicated loci in multi-ethnic cohorts from Southeast Asia. PLoS Genet 2011;7:e1001363

8. Fuchsberger C, Flannick J, Teslovich TM, et al. The genetic architecture of type 2 diabetes. Nature 2016;536:41–47

9. Auton A, Brooks LD, Durbin RM, et al.; 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature 2015;526:68–74

10. Lim ET, Würtz P, Havulinna AS, et al.; Sequencing Initiative Suomi (SISu) Project. Distribution and medical impact of loss-of-function variants in the Finnish founder population. PLoS Genet 2014;10:e1004494

11. American Diabetes Association. Classification and diagnosis of diabetes. Section 2. In *Standards of Medical Care in Diabetes—2017*. Diabetes Care 2017; 40(Suppl. 1):S11–S24

12. Kwak SH, Chae J, Choi S, et al. Findings of a 1303 Korean whole-exome sequencing study. Exp Mol Med 2017;49:e356

13. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754–1760

14. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297–1303

15. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet 2009;5:e1000529

16. Purcell S, Cherny SS, Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. Bioinformatics 2003; 19:149–150

17. Ma C, Blackwell T, Boehnke M, Scott LJ; GoT2D investigators. Recommended joint and meta-analysis strategies for case-control association testing of single low-count variants. Genet Epidemiol 2013;37:539–550

18. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26:2190–2191

19. Feng S, Liu D, Zhan X, Wing MK, Abecasis GR. RAREMETAL: fast and powerful meta-analysis for rare variants. Bioinformatics 2014;30:2828–2829

20. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38:e164

21. Madsen BE, Browning SR. A groupwise association test for rare mutations using a weighted sum statistic. PLoS Genet 2009;5:e1000384

22. Li B, Leal SM. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. Am J Hum Genet 2008;83:311–321

23. Price AL, Kryukov GV, de Bakker PI, et al. Pooled association tests for rare variants in exon-resequencing studies. Am J Hum Genet 2010;86:832–838

24. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet 2011;89:82–93

25. Purcell SM, Moran JL, Fromer M, et al. A polygenic burden of rare disruptive mutations in schizophrenia. Nature 2014;506:185–190

26. Sosa-Pineda B, Chowdhury K, Torres M, Oliver G, Gruss P. The Pax4 gene is essential for differentiation of insulin-producing beta cells in the mammalian pancreas. Nature 1997;386:399–402

27. Ritz-Laser B, Estreicher A, Gauthier BR, Mamin A, Edlund H, Philippe J. The pancreatic beta-cell-specific transcription factor Pax-4 inhibits glucagon gene expression through Pax-6. Diabetologia 2002;45:97–107

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

29. Marso SP, Daniels GH, Brown-Frandsen K, et al.; LEADER Steering Committee; LEADER Trial Investigators. Liraglutide and Cardiovascular Outcomes in Type 2 Diabetes. N Engl J Med 2016;375:311–322

30. Marso SP, Bain SC, Consoli A, et al.; SUSTAIN-6 Investigators. Semaglutide and cardiovascular outcomes in patients with type 2 diabetes. N Engl J Med 2016; 375:1834–1844

31. Jazayeri A, Rappas M, Brown AJH, et al. Crystal structure of the GLP-1 receptor bound to a peptide agonist. Nature 2017;546:254–258

32. Sathananthan A, Man CD, Micheletto F, et al. Common genetic variation in GLP1R and insulin secretion in response to exogenous GLP-1 in nondiabetic subjects: a pilot study. Diabetes Care 2010;33:2074–2076

33. Han E, Park HS, Kwon O, et al. A genetic variant in GLP1R is associated with response to DPP-4 inhibitors in patients with type 2 diabetes. Medicine (Baltimore) 2016;95:e5155

34. Kim YG, Hahn S, Oh TJ, Kwak SH, Park KS, Cho YM. Differences in the glucose-lowering efficacy of dipeptidyl peptidase-4 inhibitors between Asians and non-Asians: a systematic review and meta-analysis. Diabetologia 2013;56:696–708

35. Scott RA, Freitag DF, Li L, et al.; CVD50 consortium; GERAD_EC Consortium; Neurology Working Group of the Cohorts for Heart; Aging Research in Genomic Epidemiology (CHARGE); Alzheimer's Disease Genetics Consortium; Pancreatic Cancer Cohort Consortium; European Prospective Investigation into Cancer and Nutrition–Cardiovascular Disease (EPIC-CVD); EPIC-InterAct; CHARGE consortium; CHD Exome+ Consortium; CARDIOGRAM Exome Consortium. A genomic approach to therapeutic target validation identifies a glucose-lowering GLP1R variant protective for coronary heart disease. Sci Transl Med 2016;8: 341ra76

36. Flannick J, Thorleifsson G, Beer NL, et al.; Go-T2D Consortium; T2D-GENES Consortium. Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. Nat Genet 2014;46:357–363

37. Zhang CL, Katoh M, Shibasaki T, et al. The cAMP sensor Epac2 is a direct target of antidiabetic sulfonylurea drugs. Science 2009;325:607–610

38. Sugawara K, Shibasaki T, Takahashi H, Seino S. Structure and functional roles of Epac2 (Rapgef4). Gene 2016;575:577–583