SUPPLEMENTARY METHODS

Cohort descriptions

Discovery cohort: HAPO-HK Study

The study methods employed in the Hyperglycemia and Adverse Pregnancy Outcome-Hong Kong (HAPO-HK) Study have been previously documented [1]. The original HAPO Study is an international, prospective population-based study conducted at multiple centers. It aims to clarify the risks of adverse pregnancy outcomes linked to various degrees of maternal hyperglycemia occurring below the threshold for diagnosing overt diabetes mellitus in approximately 25,000 pregnant women during their third trimester of gestation, with medical caregivers 'blinded' to their glucose tolerance status [2]. In the HAPO-HK Study, we recruited a total of 1,674 pregnant women with singleton pregnancy at the Prince of Wales Hospital from 2000 to 2006 as part of the HAPO Study. These participants underwent a 75-g oral glucose tolerance test (OGTT) during weeks 24 to 32 of gestation, and the diagnosis of GDM was retrospectively applied based on the International Association of the Diabetes and Pregnancy Study Groups (IADPSG) criteria established in 2010. Standardized questionnaires were used to collect data on smoking and alcohol use, family history of diabetes and hypertension, as well as demographic characteristics. Individuals of non-Chinese ancestry, and those with glucose measurements outside the setting of the HAPO Study were excluded. Eligible participants were then invited to attend a follow-up assessment between 2009 and 2013, which took place approximately 7 years after childbirth. During this follow-up visit, all participants underwent a 75-g OGTT to ascertain their status of glucose tolerance, DNA samples were collected from these women. All participants provided written informed consent at the time of enrolment. The study received ethical approval (The approval numbers for the HAPO-HK study are: CRE-2002.119, CRE-2008.017, CRE-2013.042, CRE-2015.473) from the Clinical Research Ethics Committee of the Chinese University of Hong Kong.

Discovery cohort: Tianjin Study

A detailed description of the study's design, cohort, and methods was provided elsewhere [3]. Between October 2010 and August 2012, a prospective cohort study was conducted at Tianjin's Women and Children Health Center (TWCHC), recruiting a total of 22,302 pregnant women during their initial antenatal care visit through a universal screening and management system for gestational diabetes mellitus (GDM). These women participated a two-step screening procedure used to identify cases of GDM. Initially, pregnant women at 24 to 28 weeks of gestation, visiting a primary hospital, underwent a 1-hour 50-g glucose challenge test (GCT) in a non-fasting state. Women with a GCT result of ≥7.8 mmol/L were then referred to the GDM clinic at TWCHC for a 2-hour 75-g OGTT conducted in the morning after fasting for at least 8 hours. From July 2011 to June 2012, 2,991 pregnant women provided overnight fasting blood samples at the primary care hospitals. Among them, a total of 243 women with GDM and 243 women without GDM were chosen for a nested case-control study. Of them, 455 women (229 GDM cases vs. 226 non-GDM controls) with good quality DNA data were used in the current analysis. GDM was diagnosed according to the IADPSG criteria. The selection was based on matching maternal age (±1 year). The Ethics Committee for Clinical Research of Tianjin Women and Children's Health Center granted ethics approval (Ethics Approval number: 2009-02). Written informed consent was obtained from all pregnant women involved in the study.

Discovery cohort: Treated GDM Cases vs. Non-diabetes Controls (TGDM-NDM) Study

This is a case-control study that includes participants from two distinct sources. Firstly, the GDM cases were selected from an additional 141 pregnant women who were diagnosed with GDM based on OGTT at the Hong Kong center around the time of the HAPO Study. These women received antenatal treatment and were invited to return for follow-up assessment at around the same time as participants of the HAPO-HK follow-up study [4]. After excluding 47 women who were diagnosed with GDM but did not meet the criteria set by the IADPSG, along with seven women lacking genotype data and one woman with failed genotype data, a total of 86 women with GDM were included in the analysis.

Secondly, the non-diabetes and non-pregnant controls were selected from hospital staff and the "Better Health for Better Hong Kong (BHBHK)" Campaign, a community-based health awareness and promotion program [5]. The BHBHK Campaign was initiated between 2000 and 2002 and underwent re-evaluation between 2010 and 2014. Its primary objective is to raise awareness among the low-income working population about the importance of maintaining a healthy lifestyle through various educational and health screening strategies. Participants were randomly selected using stratified random sampling

based on the occupational groups documented in the 1996 Hong Kong Population By-Census Report. A total of 561 participants underwent detailed clinical and laboratory assessment at the Prince of Wales Hospital. To determine the glycemic status, these individuals underwent a 75-g OGTT using the criteria established by the American Diabetes Association. Women without diabetes at baseline and throughout the 10-year follow-up period were utilized as controls in the analysis. DNA samples from both the GDM cases and non-diabetes controls were genotyped using different genotyping platforms.

Replication cohort: Guangzhou Study

Pregnant women in their second trimester who attended the antenatal clinic of the Third Affiliated Hospital of Sun Yat-Sen University in Guangzhou between September 2013 and September 2015 for OGTT were included [6]. The prevalence of GDM during the study period was 16.8%. The diagnostic criteria for GDM were based on 75-g OGTT during weeks 24 to 28 of gestation, according to the IADPSG criteria. A total of 564 pregnant women diagnosed with GDM and 572 women without GDM were included in the study. Subjects with type 1 or type 2 diabetes mellitus, cardiovascular diseases (including hypertension and arteriosclerosis) and other metabolic disorders (including obesity body mass index [BMI] >30 kg/m^2 and polycystic ovary syndrome) were excluded. Whole blood samples collected during the OGTT were stored at –80°C prior to analysis. The study was approved by the Ethics Committee of the Third Affiliated Hospital of Sun Yat-Sen University, and all experimental procedures were performed in accordance with the relevant guidelines and regulations. Written informed consent was obtained from each individual to collect additional blood samples for genetic test.

Replication cohort: The FinnGen Project

The FinnGen Project, initiated in 2017, is a study that combines genome information with digital health care data. It operates as a public-private partnership between Finnish universities, biobanks, hospital districts, and several international pharmaceutical companies [7]. The primary objective of this project is to collect biological samples from 500,000 individuals in Finland over 6 years. Through genetic research, the project aims to enhance human health by identifying novel therapeutic targets and diagnostic tools for a wide range of diseases. The analysis in this study is based on the FinnGen release R8, which includes data from 342,499 individuals and disease endpoints. Clinical endpoints were derived from the register codes using the Finnish version of International Classification of Diseases, 10th revision (ICD-10) diagnosis codes and harmonized with definitions from ICD-8 and ICD-9. The comprehensive details on the methodology have been described elsewhere [8].

Replication cohort: The GenDIP Consortium

This study from the GENetics of Diabetes In Pregnancy (GenDIP) Consortium was a trans-ancestry meta-analysis of genetic associations for GDM in 5,485 women with GDM and 347,856 women without GDM. The descriptions for individual cohorts are available in previous reports [9].

Replication cohorts: Thai and Hispanic populations of the HAPO Study (HAPO-Thai and HAPO-Hispanic Studies)

The HAPO Study is an international epidemiological investigation with the objective of addressing unanswered questions regarding the associations between various levels of glucose intolerance and the risks of adverse outcomes during pregnancy [2]. Its general goal is to significantly advance our understanding of glucose levels during pregnancy that pose increased risks to the mother, fetus, and neonate. The study involves the collection of standardized, high-quality data on 25,000 women of diverse ethnic-racial and sociodemographic backgrounds from 16 centers worldwide. Additionally, it aims to generate data on the associations between glucose levels with the risks of adverse outcomes during pregnancy, which can be utilized to develop internationally acceptable guidelines for diagnosis and classification of GDM. For this particular study, we have exclusively included data from the Thai and Hispanic populations within the HAPO Study to validate the results of the polygenic risk score (PRS) analysis. This decision was based on the fact that only these two populations possessed complete genetic data necessary for the analysis.

The data/analyses presented in the current publication are based on the use of study data obtained from database of Genotypes and Phenotypes (dbGaP), a publicly accessible repository for genetic and phenotypic data. The specific dataset used in this study was titled "Hyperglycemia and Adverse Pregnancy Outcome (HAPO) Study: maternal glycemia and birthweight GEI Study," with accession number phs000096.v4.p1 (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000096.v4.p1). The entire dataset contains 1,500 infants and their mothers of European descent, 1,250 Af-

ro-Caribbean infants and mothers, 800 Hispanic (Mexican-American) infants and mothers, and 1,200 Thai infants and mothers. It was made available by the HAPO Study steering committee. The data download from dbGaP followed the guidelines and regulations for data access and usage as outlined by the repository.

Clinical and laboratory measurements

Descriptions of the assessment methods utilized in the HAPO-HK Study was provided elsewhere [1]. Briefly, all women underwent a standard 75-g OGTT between 24 and 32 weeks of gestation. Height, weight, and blood pressure were measured during the test visit. Data regarding smoking habits, family history of diabetes and hypertension, and demographic characteristics were collected using standardized questionnaires. Fasting blood samples were also taken to measure C-peptide levels and glycosylated hemoglobin.

Maternal weight at delivery was abstracted by research staff from the medical records. BMI was calculated as weight (kg) divided by the square of the measured height (m^2). Gestational weight gain (GWG) was computed as the mother's weight at delivery minus her pre-pregnancy weight, divided by the gestational age at delivery. Area under the curve (AUC) for glucose during OGTT at 0 to 120 minutes was calculated using the trapezoid rule. The homeostatic model assessment 2 (HOMA2) indices were computed based on the measurements of fasting glucose and C-peptide (https://www.dtu.ox.ac.uk/homacalculator/).

Genotyping, quality control, and imputation

For the genome-wide scan analysis, DNA samples from the three discovery cohorts (i.e., HAPO-HK Study, Tianjin Study, and TGDM-NDM Study) were genotyped using one of three arrays: (1) HumanOmniZhongHua-8 BeadChip, (2) Infinium Global Screening Array, and (3) Illumina Omni2.5+Exome Array. Samples from the Guangzhou Study were genotyped using the matrix-assisted laser desorption/ionization time-of-flight mass spectrometry platform.

We applied uniform quality control (QC) procedures to each set of genome-wide data. The QC of individual genotypes consisted of four steps: (1) verification of sex based on genotype calls from chromosomes X and Y; (2) identification of low-quality samples based on call rate and heterozygosity rate; (3) detection of potential familial relationships or duplicated individuals using estimates of identity-by-descent; and (4) assessment of population stratification through principal component (PC) analysis (Supplementary Fig. 5). Only biallelic autosomal single nucleotide polymorphisms (SNPs) were included in the per-marker QC. SNPs were excluded from further analysis if they met any of the following criteria: (1) Hardy–Weinberg equilibrium $P<1\times10^{-4}$; (2) minor allele frequency (MAF) <1%; or (3) call rate <95%. In particular, SNPs with MAF ≥1% but ≤5% are excluded if their call rate was <99%.

Within each cohort, the genotype data were imputed to the 1000 Genomes Project phase III reference panel (October 2014) using the Michigan Imputation Server [10]. SNPs with a MAF <1%, or an imputation quality score Rsq <0.5 were excluded. In total, 6,322,337 SNPs that were common across the discovery cohorts were included in the meta-analysis of genome-wide association studies (GWASs) for GDM.

Construction of polygenic risk score

We derived a PRS for GDM based on the variants identified in the current study. For each individual, the PRS was computed using the "score" command in PLINKv2.0 [11], represented as $\beta_1\chi_1+\beta_2\chi_2+\beta_3\chi_3+...+\beta_k\chi_k+...+\beta_n\chi_n$, where χ_k is the number of effect alleles at the k^{th} SNP, β_k is the corresponding β-coefficient for GDM risk estimated in this study, and n is the total number of SNPs involved in the PRS. The PRS was subsequently transformed into a z-score and classified into five categories using the quintile thresholds defined in the HAPO-HK Study, which is the largest cohort included in this study. Within this study, we assessed the performance of our PRS in two aspects: (1) detecting women who developed GDM during pregnancy and abnormal glucose tolerance (AGT) after pregnancy, and (2) improving the prediction of GDM and AGT after pregnancy compared to the clinical risk factors. In addition, we calculated a type 2 diabetes mellitus (T2DM)-related PRS based on 286 T2DM-related variants which were previously reported by the Diabetes Meta-Analysis of Trans-Ethnic association studies (DIAMANTE) Consortium [12]. We investigated the clinical utility of our GDM-related PRS in predicting the risk of AGT after pregnancy by comparing its performance with the T2DM-related PRS.

Statistical analyses

All analyses were performed using PLINK v1.9 and v2.0 [11], IBM SPSS Statistics version 26 (IBM Co., Armonk, NY, USA), and R 3.4.4 (http://www.r-project.org/, 31st December, 2019) unless specified otherwise. Meta-analysis was implemented by METAL software [13]. FUMA was used to annotate, prioritize,

and interpret the GWAS results [14]. Regional plot around genome-wide locus were visualized using LocusZoom (http://csg.sph.umich.edu/locuszoom/).

Data are expressed as percentage (*n*), mean and standard deviation, or median (interquartile range) as appropriate. Differences between groups were tested with chi-squared test, Student's *t*-test, or Mann-Whitney test, as appropriate.

Within the cohort, we performed logistic regression to examine the association of genetic markers (e.g., an individual SNP under an additive genetic model or the PRS) with the risks of GDM and AGT after pregnancy, with the adjustments for PCs, age and/or BMI. The results obtained from individual cohorts were combined through meta-analysis using an inverse-variance weighted approach under a fixed-effects model. Heterogeneity of effect across studies was assessed using Cochran's *Q* test. To address potential population stratification and relatedness among individuals, we adjusted for PCs in all association tests, and applied genomic control correction during the meta-analysis analysis. Associations of identified variants with glycemic and metabolic traits measured during pregnancy were tested by linear regression, adjusting for PCs, age, and/or BMI. Odds ratios with their 95% confidence intervals (CIs), or β±standard error were presented in these analyses. *P* values <0.05 and $<5.0\times10^{-8}$ were considered significant and genome-wide significant, respectively. In the candidate gene analysis, we adjusted for multiple testings using Bonferroni correction. Individuals with missing data points for any variables included in the logistic or linear regression model were removed from the analysis.

The area under the receiver operating characteristic curve (AUROC) and continuous net reclassification improvement (NRI) index were used to evaluate the incremental predictive value of PRS in GDM and AGT after pregnancy, over the clinical risk factors and PCs. We calculated the AUC and NRI index based on the predicted risk obtained from logistic regression, using respectively the "concordance.index" and "nricens" functions in R package. Bootstrapping with 10,000 iterations were used to estimate the 95% CI for the NRI index. We compared two correlated AUCs using the paired *t*-test implemented by the "cindex.comp" function in R package.

SUPPLEMENTARY RESULTS

Sensitivity analysis

In the HAPO-HK Study which included data on comprehensive clinical assessment during pregnancy, the associations between the four identified variants and GDM risk persisted after multivariate adjustment for BMI, GWG, blood pressure, smoking status, education year, parity, family history of diabetes and hypertension (Supplementary Table 14).

Associations for glycemic and metabolic traits during pregnancy

In the linear regression analysis adjusted for PCs and age, we observed several significant associations in the HAPO-HK Study: (1) the A-allele of T-box brain transcription factor 1 (*TBR1*)-solute carrier family 4 member 10 (*SLC4A10*) rs117781972 was associated with elevated levels of 1-hour glucose, 2-hour glucose and AUC_{glu} at 0 to 120 minutes ($1.7\times10^{-4}<P<0.0495$); (2) the C-allele of *CDKAL1* rs7754840 demonstrated an association with increased levels of 1-hour glucose and AUC_{glu} at 0 to 120 minutes ($0.0245<P<0.0249$); (3) there was a notable elevation in levels of 1-hour glucose, 2-hour glucose, AUC_{glu} at 0 to 120 minutes, fasting C-peptide and HOMA2 of insulin resistance (HOMA-IR) index per copy of the C-allele of *INS-IGF2*-potassium voltage-gated channel subfamily Q member 1 (*KCNQ1*) rs2237897 ($1.2\times10^{-4}<P<0.0109$); and (4) the C-allele carriers of melatonin receptor 1B (*MTNR1B*) rs7945617 had higher levels of fasting glucose, 1-hour glucose, 2-hour glucose, and AUC_{glu} at 0 to 120 minutes, as well as a reduced of HOMA2 of β-cell function (HOMA2-β) index ($1.1\times10^{-4}<P<0.0119$) (Supplementary Table 13). Adjustment for BMI did not further change these associations (Supplementary Table 13).

SUPPLEMENTARY REFERENCES

1. Tam WH, Ma RC, Ozaki R, Li AM, Chan MH, Yuen LY, et al. In utero exposure to maternal hyperglycemia increases childhood cardiometabolic risk in offspring. Diabetes Care 2017;40:679-86.
2. HAPO Study Cooperative Research Group; Metzger BE, Lowe LP, Dyer AR, Trimble ER, Chaovarindr U, et al. Hyperglycemia and adverse pregnancy outcomes. N Engl J Med 2008;358:1991-2002.
3. Leng J, Shao P, Zhang C, Tian H, Zhang F, Zhang S, et al. Prevalence of gestational diabetes mellitus and its risk factors in Chinese pregnant women: a prospective population-based study in Tianjin, China. PLoS One 2015;10:e0121029.
4. Tam WH, Ma RCW, Ozaki R, Li AM, Chan MHM, Yuen LY, et al. Antenatal treatment of gestational diabetes and offspring's