

Multi-ancestry genome-wide association study of gestational diabetes mellitus highlights genetic links with type 2 diabetes

Natalia Pervjakova^{1,†}, Gunn-Helen Moen^{2,3,4,5,†}, Maria-Carolina Borges^{5,6,†}, Teresa Ferreira^{7,†}, James P. Cook⁸, Catherine Allard⁹, Robin N. Beaumont¹⁰, Mickaël Canouil^{11,12}, Gad Hatem¹³, Anni Heiskala¹⁴, Anni Joensuu^{15,16}, Ville Karhunen^{14,17}, Soo Heon Kwak¹⁸, Frederick T.J. Lin¹⁹, Jun Liu^{20,21}, Sheryl Rifas-Shiman²², Claudia H. Tam²³, Wing Hung Tam²⁴, Gudmar Thorleifsson²⁵, Toby Andrew^{11,12,26}, Juha Auvinen¹⁴, Bishwajit Bhowmik²⁷, Amélie Bonnefond^{11,12,26}, Fabien Delahaye^{11,12}, Ayse Demirkan^{20,28}, Philippe Froguel^{11,12,26}, Kadri Haller-Kikkatalo²⁹, Hildur Hardardottir^{30,31}, Sandra Hummel^{32,33}, Akhtar Hussain^{27,34}, Eero Kajantie^{35,36,37}, Elina Keikkala^{35,36}, Amna Khamis^{11,12,26}, Jari Lahti³⁸, Tove Lekva³⁹, Sanna Mustaniemi^{35,36}, Christine Sommer⁴⁰, Aili Tagoma²⁹, Evangelia Tzala¹⁷, Raivo Uibo²⁹, Marja Väärasmäki^{36,41}, Pia M. Villa^{42,43}, Kåre I. Birkeland^{2,44}, Luigi Bouchard^{45,46}, Cornelia M. Duijn^{20,21}, Sarah Finer⁴⁷, Leif Groop^{13,48}, Esa Hämäläinen⁴⁹, Geoffrey M. Hayes^{19,50,51}, Graham A. Hitman⁴⁷, Hak C. Jang^{52,53}, Marjo-Riitta Jarvelin^{14,17}, Anne Karen Jenum⁵⁴, Hannele Laivuori^{48,55,56}, Ronald C. Ma^{23,57,58}, Olle Melander¹³, Emily Oken²², Kyong Soo Park^{18,53,59}, Patrice Perron^{9,60}, Rashmi B. Prasad¹³, Elisabeth Qvigstad^{40,61}, Sylvain Sebert¹⁴, Kari Stefansson^{25,62}, Valgerdur Steinthorsdottir²⁵, Tiinamaija Tuomi^{48,13,63,64}, Marie-France Hivert^{22,60,65}, Paul W. Franks^{66,67}, Mark I. McCarthy^{68,69,†}, Cecilia M. Lindgren^{7,69,70}, Rachel M. Freathy^{10,§}, Deborah A. Lawlor^{5,6,71,§}, Andrew P. Morris^{72,§} and Reedik Mägi^{1,§,*}

¹Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu 51010, Estonia

²Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway

³Diamantina Institute, The University of Queensland, Woolloongabba QLD 4102, Australia

⁴K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Trondheim, Norway

⁵Population Health Science, Bristol Medical School, University of Bristol, Bristol, UK

⁶MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK

⁷Big Data Institute, Li Ka Shing Center for Health for Health Information and Discovery, Oxford University, Oxford, UK

⁸Department of Health Data Science, University of Liverpool, Liverpool, UK

⁹Centre de Recherche du Centre Hospitalier Universitaire de Sherbrooke (CRCHUS), Université de Sherbrooke, Quebec, Canada

¹⁰Institute of Biomedical and Clinical Science, College of Medicine and Health, University of Exeter, Exeter, UK

¹¹Inserm U1283, CNRS UMR 8199, European Genomic Institute for Diabetes, Institut Pasteur de Lille, Lille F-59000, France

¹²University of Lille, Lille University Hospital, Lille F-59000, France

¹³Department of Clinical Sciences, Lund University, Skåne University Hospital, Lund University Diabetes Centre, Malmö SE-20502, Sweden

¹⁴Centre for Life-Course Health Research, Faculty of Medicine, University of Oulu, Oulu, Finland

¹⁵Department of Public Health Solutions, Finnish Institute for Health and Welfare, Helsinki, Finland

¹⁶Research Program for Clinical and Molecular Metabolism, Faculty of Medicine, University of Helsinki, Helsinki, Finland

¹⁷School of Public Health, Department of Epidemiology and Biostatistics, Imperial College London, St Mary's Hospital, London, UK

¹⁸Department of Internal Medicine, Seoul National University Hospital, Seoul, Republic of Korea

¹⁹Division of Endocrinology, Metabolism, and Molecular Medicine, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA

²⁰Department of Epidemiology, Erasmus Medical Center, Rotterdam, the Netherlands

²¹Nuffield Department of Population Health, University of Oxford, Oxford, UK

²²Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA, USA

²³Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong SAR, The People's Republic of China

²⁴Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong, Hong Kong SAR, The People's Republic of China

²⁵deCODE Genetics/Amgen, Inc., Reykjavik, Iceland

²⁶Department of Metabolism, Digestion and Reproduction, Imperial College London, London, UK

²⁷Centre of Global Health Research, Diabetic Association of Bangladesh, Dhaka, Bangladesh

²⁸Section of Statistical Multi-omics, Department of Clinical and Experimental Research, University of Surrey, Surrey, UK

²⁹Department of Immunology, Institute of Biomedicine and Translational Medicine, University of Tartu, Tartu, Estonia

³⁰Faculty of Medicine, University of Iceland, Reykjavik, Iceland

³¹Livio Reykjavik, Reykjavik, Iceland

Received: November 11, 2021. Revised: February 9, 2022. Accepted: February 23, 2022

© The Author(s) 2022. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

- ³²Institute of Diabetes Research, Helmholtz Zentrum München, German Research Center for Environmental Health, Munich-Neuherberg, Germany
- ³³Forschergruppe Diabetes, Technical University Munich, at Klinikum rechts der Isar, Munich, Germany
- ³⁴Faculty of Health Sciences, Nord University, Bodø, Norway
- ³⁵Population Health Unit, Finnish Institute for Health and Welfare, Helsinki and Oulu, Finland
- ³⁶PEDEGO Research Unit, MRC Oulu, Oulu University Hospital and University of Oulu, Oulu, Finland
- ³⁷Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway
- ³⁸Department of Psychology and Logopedics, University of Helsinki, Helsinki, Finland
- ³⁹Research Institute of Internal Medicine, Oslo University Hospital, Oslo, Norway
- ⁴⁰Department of Endocrinology, Morbid Obesity and Preventive Medicine, Oslo University Hospital, Oslo, Norway
- ⁴¹Population Health Unit, Finnish Institute for Health and Welfare, Helsinki and Oulu, Finland
- ⁴²Department of Obstetrics and Gynaecology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland
- ⁴³Hyvinkää Hospital, Helsinki and Uusimaa Hospital District, Hyvinkää, Finland
- ⁴⁴Department of Transplantation Medicine, Oslo University Hospital, Oslo, Norway
- ⁴⁵Department of Biochemistry and Functional Genomics, Faculty of Medicine and Health Sciences, Université de Sherbrooke, Québec, Canada
- ⁴⁶Department of Medical Biology, Centre Intégré Universitaire de Santé et de Services Sociaux du Saguenay–Lac-St-Jean – Hôpital de Chicoutimi, Québec, Canada
- ⁴⁷Centre for Genomics and Child Health, Blizard Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, UK
- ⁴⁸Institute for Molecular Medicine Finland (FIMM), Helsinki Institute of Life Science, University of Helsinki, Helsinki, Finland
- ⁴⁹Department of Clinical Chemistry, University of Eastern Finland, Kuopio, Finland
- ⁵⁰Center for Genetic Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA
- ⁵¹Department of Anthropology, Northwestern University, Evanston, IL 60208, USA
- ⁵²Department of Internal Medicine, Seoul National University Bundang Hospital, Seongnam, Republic of Korea
- ⁵³Department of Internal Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea
- ⁵⁴General Practice Research Unit (AFE), Department of General Practice, Institute of Health and Society, Faculty of Medicine, University of Oslo, Post Box 1130 Blindern, Oslo 0318, Norway
- ⁵⁵Department of Obstetrics and Gynecology, Tampere University, Hospital and Faculty of Medicine and Health Technology, Center for Child, Adolescent, and Maternal Health, Tampere University, Tampere, Finland
- ⁵⁶Medical and Clinical Genetics, University of Helsinki and Helsinki University Hospital, Helsinki, Finland
- ⁵⁷Laboratory for Molecular Epidemiology in Diabetes, Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Hong Kong SAR, The People's Republic of China
- ⁵⁸Hong Kong Institute of Diabetes and Obesity, The Chinese University of Hong Kong, Hong Kong SAR, The People's Republic of China
- ⁵⁹Department of Molecular Medicine and Biopharmaceutical Sciences, Graduate School of Convergence Science and Technology, Seoul National University, Seoul, Republic of Korea
- ⁶⁰Department of Medicine, Faculty of Medicine and Health Sciences, University of Sherbrooke, Québec, Canada
- ⁶¹Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway
- ⁶²Faculty of Medicine, University of Iceland, Reykjavik, Iceland
- ⁶³Department of Endocrinology, Abdominal Centre, Helsinki University Hospital, Helsinki, Finland
- ⁶⁴Folkhalsan Research Center, Helsinki, Finland
- ⁶⁵Diabetes Unit, Massachusetts General Hospital, Boston, MA, USA
- ⁶⁶Department of Clinical Sciences, Lund University, Malmö, Sweden
- ⁶⁷Department of Nutrition, Harvard School of Public Health, Boston, MA, USA
- ⁶⁸Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, UK
- ⁶⁹Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK
- ⁷⁰Program in Medical and Population Genetics, Broad Institute, Boston, MA, USA
- ⁷¹Bristol NIHR Biomedical Research Centre, Bristol, UK
- ⁷²Centre for Genetics and Genomics Versus Arthritis, Centre for Musculoskeletal Research, Division of Musculoskeletal and Dermatological Sciences, University of Manchester, Manchester, UK
- *To whom correspondence should be addressed at: Institute of Genomics, University of Tartu, Riia 23B/2, Tartu 51010, Estonia. Tel: +372 7374045; Fax: +372 7374060; Email: reedik.magi@ut.ee

[†]Present address: Genentech, 1 DNA Way, South San Francisco, CA, USA.

[‡]These authors contributed equally to this work.

[§]These authors supervised this work.

Abstract

Gestational diabetes mellitus (GDM) is associated with increased risk of pregnancy complications and adverse perinatal outcomes. GDM often reoccurs and is associated with increased risk of subsequent diagnosis of type 2 diabetes (T2D). To improve our understanding of the aetiological factors and molecular processes driving the occurrence of GDM, including the extent to which these overlap with T2D pathophysiology, the GENetics of Diabetes In Pregnancy Consortium assembled genome-wide association studies of diverse ancestry in a total of 5485 women with GDM and 347 856 without GDM. Through multi-ancestry meta-analysis, we identified five loci with genome-wide significant association ($P < 5 \times 10^{-8}$) with GDM, mapping to/near *MTNR1B* ($P = 4.3 \times 10^{-54}$), *TCF7L2* ($P = 4.0 \times 10^{-16}$), *CDKAL1* ($P = 1.6 \times 10^{-14}$), *CDKN2A-CDKN2B* ($P = 4.1 \times 10^{-9}$) and *HKDC1* ($P = 2.9 \times 10^{-8}$). Multiple lines of evidence pointed to the shared pathophysiology of GDM and T2D: (i) four of the five GDM loci (not *HKDC1*) have been previously reported at genome-wide significance for T2D; (ii) significant enrichment for associations with GDM at previously reported T2D loci; (iii) strong genetic correlation between GDM and T2D and (iv) enrichment of GDM associations mapping to genomic annotations in diabetes-relevant tissues and transcription factor binding sites. Mendelian randomization analyses demonstrated significant causal association (5% false discovery rate) of higher body mass index on increased GDM risk. Our results provide support for the hypothesis that GDM and T2D are part of the same underlying pathology but that, as exemplified by the *HKDC1* locus, there are genetic determinants of GDM that are specific to glucose regulation in pregnancy.

Introduction

Gestational diabetes mellitus (GDM), defined as hyperglycaemia with onset or first recognition during pregnancy, is associated with increased risk of pregnancy complications and adverse perinatal outcomes, including pre-eclampsia, stillbirth, large for gestational age, neonatal hypoglycaemia, preterm birth, low Apgar scores and admission to neonatal intensive care (1–4). Whilst hyperglycaemia commonly resolves postpartum, GDM often reoccurs (5) and is associated with subsequent diagnosis of type 2 diabetes (T2D) and coronary heart disease (6,7). Although the global prevalence of GDM is increasing, it varies according to population characteristics (such as maternal age, ancestry and obesity rates) and the criteria used for screening and diagnosis (8).

GDM and T2D share both genetic and non-genetic risk factors, including obesity, poor diet and sedentary lifestyle (9,10). Family studies have demonstrated that women with GDM have 30.1% probability of having at least one parent with T2D, compared to just 13.2% for pregnant women with normal glucose tolerance (11). Furthermore, women with a history of GDM appear to have a nearly 10-fold higher risk of developing T2D than those with a normoglycaemic pregnancy (7). Taken together, these observations support the hypothesis that the two diseases are part of the same underlying pathology, with pregnancy potentially acting as a stress test that reveals women at increased risk of GDM and/or T2D (12,13).

There have been considerable advances in our understanding of the genetic contribution to T2D through large-scale genome-wide association studies (GWAS) across diverse populations (14–17). In contrast, despite the observed familial clustering of GDM (18), most genetic association studies of the disease have focussed on evaluating the impact of previously reported loci for T2D and glycaemic traits in modest sample sizes (19). The most comprehensive systematic review of genetic susceptibility to GDM (from 23 studies) revealed association with T2D risk variants from seven loci, of which six are related to insulin secretion and one to insulin resistance (20). A genetic risk score (GRS) of risk variants across 34 loci associated with T2D and/or fasting glucose was significantly associated with GDM and improved predictive power over a model including only clinical variables (21). Variants associated with both insulin secretion and insulin resistance have also been used to construct an aggregated GRS that was shown to predict GDM risk, with and without adjustment for body mass index (BMI), maternal age and gestational age, although this score was not compared with established clinical predictors (22). To date, the largest GWAS of GDM has been undertaken in women from a Korean population, including 468 cases and 1242 non-diabetic controls in the discovery stage, with an additional 931 cases and 783 non-diabetic controls in the follow-up stage (23). Two loci were associated with GDM at genome-wide significance ($P < 5 \times 10^{-8}$), mapping near *MTNR1B*

and *CDKAL1*, both of which have also been previously implicated in T2D risk.

To gain novel insight into the genetic architecture of GDM, the GENetics of Diabetes In Pregnancy (GenDIP) Consortium assembled GWAS of diverse ancestry in a total of 5485 women with GDM and 347 856 women without GDM. With these resources, we aimed to improve our understanding of the aetiological factors and molecular processes driving the occurrence of GDM, including the extent to which these overlap with T2D pathophysiology, and investigate the effects of potential causal metabolic risk factors on the disease through Mendelian randomization (MR).

Results

We began by aggregating GDM association summary statistics across GWAS through multi-ancestry meta-analysis: the effective sample size was 72.2% European, 13.4% East Asian, 9.9% South Asian, 2.8% Hispanic/Latino and 1.7% African (Supplementary Material, Tables S1 and S2). To maximize sample size, we used a phenotype definition that makes best use of the information available in each study, including data from health records, oral glucose tolerance tests and self-report (Supplementary Material, Table S1). Each GWAS was imputed to reference panels from the 1000 Genomes Project (24), Haplotype Reference Consortium (25) or population-specific whole-genome sequence data (Supplementary Material, Table S3). Within each GWAS, GDM association summary statistics were derived for all single nucleotide variants (SNVs) passing quality control after appropriate adjustment to account for population structure (Supplementary Material, Table S3).

The most powerful methods for multi-ancestry meta-analysis allow for potential allelic effect heterogeneity on disease between population groups that cannot be accommodated in a fixed-effects model (26). Our primary analysis used MR-MEGA (27), which models heterogeneity between GWAS by including axes of genetic variation that represent ancestry as covariates in a meta-regression model. We considered three axes of genetic variation that separated the five ancestry groups, but which also revealed finer-scale genetic differences between GWAS of the same ancestry (Supplementary Material, Fig. S1). We also conducted multi-ancestry and ancestry-specific fixed-effects meta-analyses. We identified five loci at genome-wide significance in the multi-ancestry meta-regression (Table 1, Supplementary Material, Figs S2 and S3), including the previously reported associations from GDM GWAS at *MTNR1B* (rs10830963, $P = 4.3 \times 10^{-54}$) and *CDKAL1* (rs9348441, $P = 1.6 \times 10^{-14}$). The remaining three loci for GDM mapped to/near *TCF7L2* (rs7903146, $P = 4.0 \times 10^{-16}$), *CDKN2A-CDKN2B* (rs10811662, $P = 4.1 \times 10^{-9}$) and *HKDC1* (rs9663238, $P = 2.9 \times 10^{-8}$). Through approximate conditional analyses, conducted using ancestry-matched

Table 1. Loci attaining genome-wide significant ($P < 5 \times 10^{-8}$) evidence of association with GDM in multi-ancestry meta-regression (MR-MEGA) of 5485 cases and 347 856 controls

| Locus | Lead SNV | Chr | Position (bp, b37) | Alleles | | MR-MEGA P-value | Fixed-effects OR (95% CI) |
|---------------|------------|-----|--------------------|---------|-------|-----------------------|------------------------------|
| | | | | Risk | Other | | |
| MTNR1B | rs10830963 | 11 | 92 708 710 | G | C | 4.3×10^{-54} | 1.41 (1.35–1.47) |
| TCF7L2 | rs7903146 | 10 | 114 758 349 | T | C | 4.0×10^{-16} | 1.22 (1.16–1.27) |
| CDKAL1 | rs9348441 | 6 | 20 680 678 | A | T | 1.6×10^{-14} | 1.13 (1.08–1.18) |
| CDKN2A-CDKN2B | rs10811662 | 9 | 22 134 253 | G | A | 4.1×10^{-9} | 1.14 (1.09–1.20) |
| HKDC1 | rs9663238 | 10 | 70 983 629 | G | A | 2.9×10^{-8} | 1.14 (1.09–1.19) |

Chr: chromosome. OR: odds-ratio. CI: confidence interval.

linkage disequilibrium (LD) reference panels for each GWAS, we observed no evidence for multiple distinct association signals at genome-wide significance at any of the five GDM loci (Supplementary Material, Fig. S4).

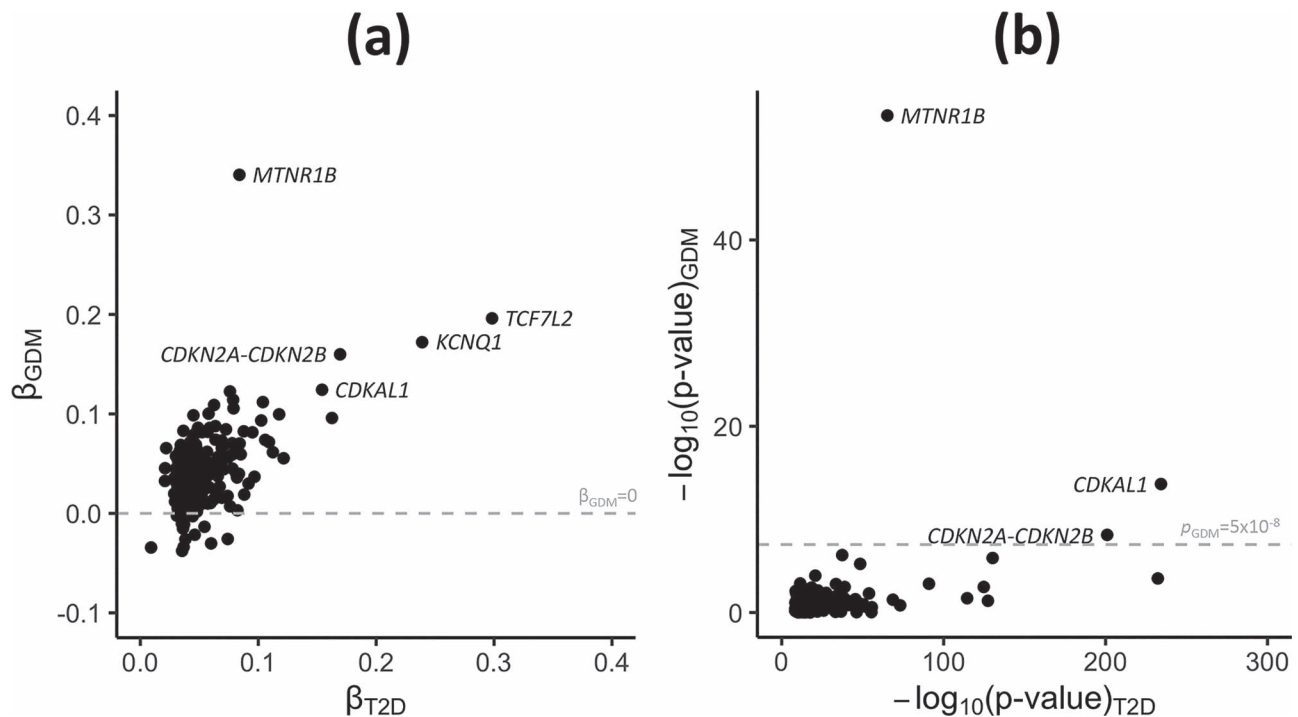
We next sought to investigate the impact of differences in ancestry and phenotype definition between GWAS on heterogeneity in allelic effects at GDM loci. To do this, we extended the MR-MEGA meta-regression model to include an additional covariate to represent whether GDM status in the study was confirmed via 'a universal blood-based test' (Supplementary Material, Table S1). Here, we use this term to refer to a blood-based test that was applied to all participants, including a diagnostic oral glucose tolerance test (OGTT) or a screening glucose challenge or fasting glucose test, in contrast to clinician decision, risk factor screening or a lack of clarity on what basis women did or did not have a diagnostic OGTT. This model enables partitioning of heterogeneity into three components (Table 2). The first component captures heterogeneity that is correlated with genetic ancestry (that can be explained by the three axes of genetic variation), which can occur because of differences in the structure of LD between ancestry groups or interactions with lifestyle factors that vary across populations. The second component measures heterogeneity that can be explained by the use of a universal blood-based test to screen for or diagnose GDM. The final component reflects residual heterogeneity due to study design that cannot be explained by the first two components. The greatest evidence of ancestry-correlated heterogeneity (after accounting for the use of a universal blood-based test) was observed at the CDKAL1 locus ($P_{\text{HET}} = 3.4 \times 10^{-5}$), where the lead SNV demonstrated stronger effects on GDM in GWAS of East Asian ancestry than in other populations, despite the risk allele being common in all ancestry groups (Supplementary Material, Fig. S5, Table S4). A similar pattern of ancestry-correlated heterogeneity in allelic effects on T2D has been reported at the CDKAL1 locus (16). Weaker evidence of ancestry-correlated heterogeneity was observed at the CDKN2A-CDKN2B locus ($P_{\text{HET}} = 0.0022$), where there were marked differences in the effects on GDM of the lead SNV between GWAS undertaken in different ancestry groups (Supplementary Material, Fig. S5, Table S4). In contrast, there was no evidence of heterogeneity due to phenotype definition for

any lead SNV, suggesting that differences in allelic effects between GWAS are more likely due to factors related to genetic ancestry than the use of a blood-based test in all women to screen for or diagnose GDM.

Of the five GDM loci identified at genome-wide significance in the trans-ancestry meta-regression, four have been previously implicated in T2D susceptibility: MTNR1B, TCF7L2, CDKAL1 and CDKN2A-CDKN2B. In fact, in previously reported trans-ancestry GWAS meta-analyses of 180 834 T2D cases and 1 159 055 controls from the DIAMANTE Consortium (16), the lead T2D SNV is the same as we report for GDM at MTNR1B, TCF7L2 and CDKAL1, and is in strong linkage disequilibrium (LD) at the CDKN2A-CDKN2B locus (rs10811661, $r^2 = 0.91$ across diverse populations in the 1000 Genomes Project (24)). To further investigate the genetic correlation between the two diseases, we extracted GDM association summary statistics from our trans-ancestry meta-analysis for lead SNVs at 222 previously reported loci for T2D from the DIAMANTE Consortium (16) (Supplementary Material, Fig. 1, Table S5). We observed a strong positive correlation in log-ORs for the T2D risk allele between the two diseases: Pearson $r = 0.573$ ($P < 2.2 \times 10^{-16}$). There was also a highly significant enrichment of GDM associations at T2D loci (50 of 222 lead SNVs with $P < 0.05$ and same direction of effect, binomial test $P < 2.2 \times 10^{-16}$), indicating that they would be discovered at genome-wide significance with larger effective sample sizes. Indeed, after excluding the four overlapping GDM-T2D loci, a weighted genetic risk score of lead T2D SNVs was significantly associated with GDM ($P = 9.7 \times 10^{-123}$, pseudo- $R^2 = 2.86\%$). Extending our analyses, genome-wide, using LD-score regression, we observed strong genetic correlation between GDM and T2D: r_G (95% CI) 0.744 (0.052, 1.437). Weaker genetic correlations between GDM and other glycaemic traits, BMI and birth weight were also observed (Table 3, Supplementary Material, Table S6). These results are consistent with sharing of genetic determinants of GDM and T2D, although we acknowledge that LD score regression has limited statistical power because of the relatively small GDM sample size, and we note that the correlation from LD-score regression is not bound by -1 to 1 , particularly when power is low.

Table 2. Source of heterogeneity in allelic effects on GDM between GWAS for lead SNVs derived from meta-regression of 5485 cases and 347 856 controls

| Locus | Lead SNV | Heterogeneity source (P-value) | | |
|---------------|------------|--------------------------------|----------------------------|----------------------|
| | | Ancestry | Universal blood-based test | Residual |
| MTNR1B | rs10830963 | 0.14 | 0.41 | 0.67 |
| TCF7L2 | rs7903146 | 0.25 | 0.83 | 0.089 |
| CDKAL1 | rs9348441 | 3.4×10^{-5} | 0.28 | 0.15 |
| CDKN2A-CDKN2B | rs10811662 | 0.0022 | 0.45 | 0.26 |
| HKDC1 | rs9663238 | 0.19 | 0.33 | 5.4×10^{-5} |

**Figure 1.** Correlation between GDM and T2D association summary statistics for lead SNVs at previously reported loci for T2D susceptibility. Association summary statistics for GDM were obtained from multi-ancestry GWAS meta-analyses of 5485 cases and 347 856 controls. Association summary statistics for T2D were obtained from multi-ancestry GWAS meta-analyses of 180 834 cases and 1 159 055 controls from the DIAMANTE Consortium. (a) Allelic effect sizes (log-ORs) for each disease, aligned to the T2D risk allele, from fixed-effects meta-analysis. The grey line represents log-OR of zero for GDM. (b) Association evidence ($-\log_{10}$ P-values) for each disease from meta-regression. The grey line represents genome-wide significance ($P < 5 \times 10^{-8}$) for GDM. The lead SNV at the TCF7L2 locus has been removed for ease of presentation (Supplementary Material, Table S4).

The most obvious difference in allelic effect sizes between GDM and T2D was observed at the *MTNR1B* locus (Fig. 1). The lead SNV, rs10830963, is the same for both diseases, but the allelic effect on GDM is substantially greater than on T2D: OR (95% CI) for GDM is 1.41 (1.35–1.47) and for T2D is just 1.09 (1.08–1.10). The *MTNR1B* lead SNV is associated, at genome-wide significance, with fasting glycaemic traits in non-diabetic individuals from the Meta-Analysis of Glucose and Insulin-related traits Consortium (MAGIC) Investigators (28,29). The GDM risk allele at the lead SNV is also associated with higher fasting plasma glucose and 1-hour plasma glucose in pregnant women from the Hyperglycemia and Adverse Pregnancy Outcomes (HAPO) Study (30). This SNV also has the strongest association of the maternal glucose-raising allele with higher offspring birth weight

in women from the Early Growth Genetics Consortium (31), in line with the known effects of maternal hyperglycaemia on fetal growth. In non-diabetic individuals from the MAGIC Investigators (32), the *MTNR1B* lead SNV has a much larger impact on fasting glucose than those at *TCF7L2*, *CDKAL1* and *CDKN2A-CDKN2B* (33) (Supplementary Material, Table S7). Therefore, the difference in allelic effect sizes between GDM and T2D at *MTNR1B* may reflect the fact that thresholds of fasting plasma glucose used to diagnose GDM are lower than those used to diagnose T2D, meaning that a larger proportion of GDM than T2D cases will have higher fasting glucose that is regulated within the normal range.

To gain insight into the molecular processes and tissues through which GDM association signals are mediated, genome-wide, we then undertook fGWAS

Table 3. Genetic correlation from LD-score regression of GDM with T2D and glycaemic traits, BMI and birth weight

| Trait | Genetic correlation r_G (95% CI) ^a |
|--------------------------------|---|
| T2D | 0.744 (0.052, 1.437) |
| Fasting glucose | 0.218 (−0.211, 0.648) |
| Fasting insulin | 0.410 (−0.114, 0.934) |
| Fasting proinsulin | 0.336 (−0.321, 0.993) |
| 2 h glucose (adjusted for BMI) | 0.444 (−0.371, 1.260) |
| HbA1c | 0.387 (−0.218, 0.991) |
| HOMA-B | −0.005 (−0.551, 0.541) |
| HOMA-IR | 0.236 (−0.382, 0.854) |
| BMI | 0.405 (0.001, 0.809) |
| Birthweight (maternal) | −0.085 (−0.358, 0.189) |
| Birthweight (fetal) | −0.059 (−0.295, 0.178) |

CI: Confidence Interval. T2D: type 2 diabetes. HbA1c: glycated haemoglobin. BMI: body mass index. HOMA-B: homeostasis model assessment of β -cell activity; HOMA-IR: homeostasis model assessment of insulin resistance. ^aGenetic correlation obtained from LD-score regression is not bound by -1 to 1 and estimates can therefore be found outside these limits due to high imprecision caused by factors such as low sample size in the association summary statistics used.

enrichment analyses within three categories of functional and regulatory annotations: (i) genic regions (34); (ii) chromatin immuno-precipitation sequence (ChIP-seq) binding sites for 165 transcription factors (35,36) and (iii) 13 unique and recurrent chromatin states in four diabetes-relevant tissues (pancreatic islets, liver, adipose, and skeletal muscle) (37). We observed significant joint enrichment ($P < 0.05$) for GDM associations mapping to protein coding exons, binding sites for FOXA2, NFE2 and TFAP2, and chromatin states in adipose tissue and skeletal muscle that mark enhancers and transcribed regions (Supplementary Material, Table S8). FOXA2 is a pioneer factor involved in pancreatic and hepatic development, and T2D association signals have been previously reported to be enriched for FOXA2-binding sites (38). Skeletal muscle is the most prominent site of insulin-mediated glucose uptake in humans, and enhancers in skeletal muscle have been reported to overlap association signals for metabolic disorders, including T2D, insulin resistance and obesity (39). These enrichment analyses highlight molecular processes and tissues that are broadly consistent with those important in mediating T2D association signals (16), although the involvement of pancreatic islets appears to be less prominent for GDM.

In contrast to the other GDM loci reported in this investigation, the lead SNV at the *HKDC1* locus (rs9663238) demonstrates only weak statistical evidence of T2D association in previously reported trans-ancestry GWAS meta-analyses from the DIAMANTE Consortium (16) ($P = 0.0083$, compared with $P < 10^{-65}$ at the other four loci). GDM risk alleles at the lead SNV, and/or at variants in strong LD (European ancestry $r^2 > 0.9$) with it, have been previously associated, at genome-wide significance, with higher 2-h plasma glucose (2HPG) in pregnant women in the HAPO Study and two replication studies of European ancestry (30), as well as with higher birth weight of first child (likely via greater maternal

glucose availability), higher own birth weight (fetal effect independent of the maternal effect on birth weight) and comparative height and body size at age 10 in UK Biobank (40,41) (Supplementary Material, Table S9). The lead SNV is also associated, more strongly in women than men, with higher alanine aminotransferase (ALT) levels in UK Biobank (42). Elevated ALT levels in early pregnancy have been associated with the risk of subsequent development of GDM (43) and genome-wide, we observed positive genetic correlation between the two traits: r_G (95% CI) 0.149 (0.005, 0.292). In addition to demonstrating the association of the maternal SNVs at this locus with GDM in the current study, we observed that 99% credible set variants are lead SNVs for *HKDC1* expression quantitative trait loci in a range of tissues in the GTEx Project (44), including visceral adipose, subcutaneous adipose and pancreas (Supplementary Material, Table S10). *HKDC1* (Hexokinase Domain Containing 1) catalyses the phosphorylation of hexose to hexose 6-phosphate and is involved in glucose homeostasis and hepatic lipid accumulation. Haplotypes of variants associated with 2HPG in pregnancy disrupt regulatory element activity and reduce *HKDC1* expression across diverse tissues (including metabolically relevant liver stellate cells and pancreatic islet beta cells), which has been demonstrated to reduce hexokinase activity in multiple cellular models (45). Knockout of hepatic *HKDC1* in pregnant mice has also been demonstrated to significantly impair glucose tolerance, highlighting the importance of liver *HKDC1* on glucose metabolism during pregnancy (46). Taken together, the evidence from our study and others suggests a more important role for *HKDC1* in glucose metabolism during pregnancy than outside of pregnancy, in addition to independent maternal and offspring effects on early growth, and highlights that while GDM shares many similarities with T2D, there are differences in at least one underlying pathway.

Finally, we used two-sample MR to investigate causal effects on GDM of 282 metabolic measures and risk factors available in the MR-Base GWAS catalogue (www.mrbase.org) (47), including metabolites, anthropometric measures, hormones, immune system phenotypes, kidney traits and metals (Supplementary Material, Table S11). We did not consider glycaemic traits (including HbA1c) because they are used to define GDM status. For each metabolic measure, we selected independent SNVs attaining genome-wide significance with the trait as instrumental variables. For each SNV, we extracted association summary statistics for GDM from the European ancestry-specific meta-analysis because we assessed independence of genetic instruments using LD from European ancestry haplotypes from the 1000 Genomes Project (24). Of the 282 exposures considered, only BMI demonstrated significant evidence for a causal effect on GDM risk at a false discovery rate of 5% (Supplementary Material, Table S11). The estimated causal effect of higher BMI on higher GDM risk was directionally consistent across multiple MR models

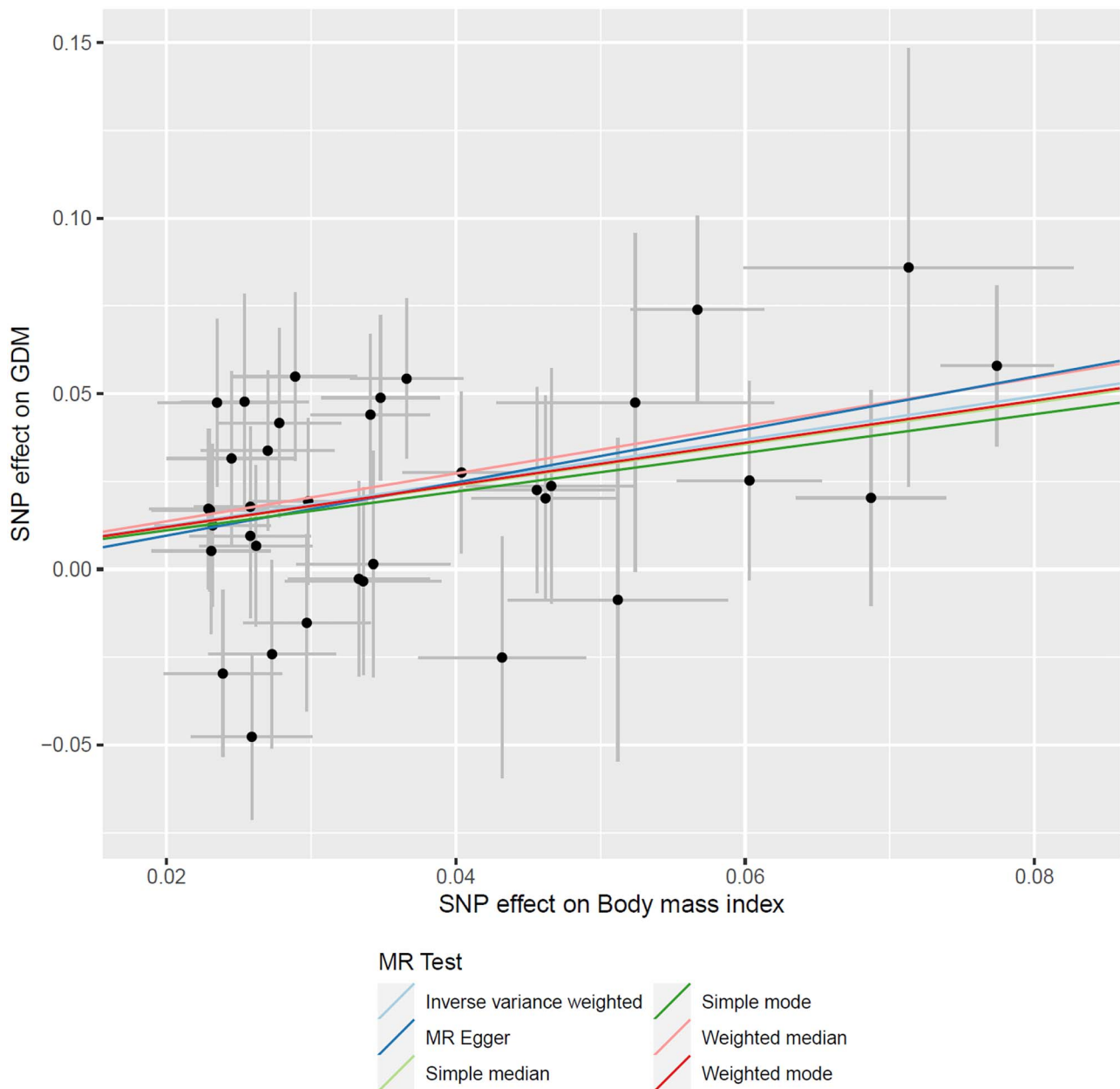


Figure 2. Effects of BMI on GDM from MR analyses. Each point corresponds to an independent SNV (genetic instrument), plotted according to the effect on BMI (on the x-axis) and the effect on GDM (log-OR, on the y-axis). Horizontal and vertical bars represent the standard errors of effect estimates. The coloured regression lines represent the effect of BMI on GDM from six MR models.

(Fig. 2). The causal relationship of BMI with GDM is consistent with its effect on T2D (48).

Discussion

We have conducted the largest and most ancestrally diverse GWAS meta-analysis for GDM, where we identified associations mapping to *MTNR1B*, *TCF7L2*, *CDKAL1*, *CDKN2A-CDKN2B* and *HKDC1*. Our results demonstrated strong correlation in the effects of previously reported associations for T2D and those observed for GDM, and highlighted overlapping molecular mechanisms and tissues that mediate associations for both diseases. In contrast, variation at the *HKDC1* locus is not strongly associated with T2D, but instead plays a more important

role in glucose metabolism during pregnancy than outside of pregnancy. The genetic diversity of GWAS contributing to our meta-analysis enabled identification of ancestry-correlated heterogeneity in allelic effects on GDM at two loci. Such heterogeneity could reflect variable impact of different pathophysiology driving glycaemic dysregulation in pregnancy between ancestries and emphasizes the need for increased sample sizes in under-represented population groups. In contrast, results were consistent between GWAS in which all women had a universal blood-based test and those that did not, suggesting little impact from misclassification due to selective use of diagnostic tests only in those deemed to be at high-risk. Finally, MR analyses revealed a significant causal effect of higher BMI on GDM risk, consistent

with the causal association observed with T2D. Taken together, these results provide further support for the hypothesis that T2D and GDM are part of the same underlying pathology. However, they also highlight there are pathways to GDM that impact on glucose regulation only in pregnancy, and that additional GDM-specific associations will be revealed through GWAS in larger sample sizes.

Materials and Methods

Ethics statement

All human research was approved by the relevant institutional review boards and conducted according to the Declaration of Helsinki. All participants provided written informed consent.

Study-level analyses

Individuals were assayed with a range of GWAS genotyping arrays, with sample and SNV quality control undertaken within each study (Supplementary Material, Tables S2 and S3). Samples were pre-phased and imputed up to reference panels from the 1000 Genomes Project (phase 1, March 2012 release; phase 3, October 2014 release) (24,49), Haplotype Reference Consortium (25) or population-specific whole-genome sequencing (50–52) (Supplementary Material, Table S3). SNVs with poor imputation quality ($r^2 < 0.3$ or $\text{info} < 0.4$) and/or minor allele count < 5 were excluded from downstream association analyses (Supplementary Material, Table S3). Association with GDM was evaluated in a regression framework, under an additive model in the dosage of the minor allele, with adjustment for principal components and other study-specific covariates to minimize the population stratification effects (Supplementary Material, Table S3). Phenotype definition and covariate adjustments were not harmonized between GWAS because of differences in individual study design and availability of non-genetic risk factor information. Analyses accounted for structure (population stratification and/or familial relationships) by: (i) excluding related samples and adjustment for principal components derived from a genetic relatedness matrix (GRM) as additional covariates in the regression model; or (ii) incorporating a random effect for the GRM in a mixed model (Supplementary Material, Table S3). Allelic effects and corresponding standard errors that were estimated from a linear (mixed) model were converted to the log-odds scale (53). Study-level association summary statistics (P -values and standard error of allelic effects) were corrected for residual structure by means of genomic control (54) if the inflation factor was > 1 (Supplementary Material, Table S3).

Multi-ancestry meta-analyses

To account for the different reference panels used for imputation across GWAS, we restricted our analyses to autosomal bi-allelic SNVs from the 1000 Genomes

Project reference panel (phase 3, October 2014 release) (24) that are also present in the Haplotype Reference Consortium reference panel (25). We considered only those SNVs with $\text{MAF} > 0.5\%$ in haplotypes in at least one of the five ancestry groups represented in the 1000 Genomes Project (phase 3, October 2014 release).

Our primary multi-ancestry analysis utilized meta-regression, implemented in the MR-MEGA software, which allows for allelic effect heterogeneity between GWAS that is correlated with ancestry (27). We first constructed a distance matrix of mean effect allele frequency differences between each pair of GWAS across a subset of SNVs reported in all studies. We implemented multi-dimensional scaling of the distance matrix to obtain three principal components that defined axes of genetic variation to separate GWAS from the five ancestry groups (Supplementary Material, Fig. S1). For each SNV, we then modelled allelic log-ORs across GWAS in a linear regression framework, weighted by the inverse of the variance of the effect estimates, incorporating the three axes of genetic variation as covariates. Under this model, we tested for association with GDM allowing for allelic effect heterogeneity between GWAS that is correlated with ancestry. We corrected the meta-regression association P -values for inflation due to residual structure between GWAS using genomic control adjustment. We considered only those SNVs reported $\geq 50\%$ of the total effective sample size in downstream analyses.

For each SNV, we also conducted fixed-effects meta-analysis across GWAS under an inverse-variance weighting of allelic log-ORs using GWAMA (55). We corrected standard errors of the resulting effect estimates for inflation due to residual structure between GWAS by genomic control adjustment.

Defining GDM loci

We identified lead SNVs attaining genome-wide significant evidence of association ($P < 5 \times 10^{-8}$) in the multi-ancestry meta-regression that were separated by at least 500 kb. Loci were defined by the genomic interval mapping 500 kb up- and downstream of each lead SNV.

Assessing evidence for multiple distinct association signals at GDM loci

Each GWAS was first assigned to one of the ancestry groups (Supplementary Material, Table S2) represented in the 1000 Genomes Project reference panel (phase 3, October 2014 release) (24). Haplotypes in the panel that were specific to that ancestry group were used as a reference for LD between SNVs across loci for the GWAS in approximate conditional analyses implemented in GCTA (56). For each locus, we applied GCTA to each GWAS to condition on the lead SNV at the locus, using the study-level association summary statistics and matched LD reference. Allelic log-ORs from the approximate conditional analyses across GWAS were modelled in the multi-ancestry meta-regression framework implemented

in MR-MEGA (27), incorporating the three axes of genetic variation as covariates, and weighted by the inverse of the variance of the effect estimates. The meta-regression association *P*-values were corrected for inflation due to residual structure between GWAS by using the same genomic control adjustment as in the unconditional analysis. If no SNVs attained genome-wide significant ($P < 5 \times 10^{-8}$) evidence of residual GDM association in the meta-regression, we concluded that there was a single association signal at the locus.

Ancestry-specific meta-analyses

We aggregated association summary statistics across GWAS from the same ancestry group via fixed-effects meta-analysis based on inverse-variance weighting of allelic log-OR to obtain effect size estimates using GWAMA (55). We corrected association *P*-values and standard errors of allelic effects from each ancestry group for residual inflation due to structure between GWAS by genomic control adjustment if the inflation factor was >1 . We estimated the mean effect allele frequency across GWAS from each ancestry group, weighted by the effective sample size of the study.

Investigating the source of heterogeneity in allelic effects on GDM

We extended the meta-regression model implemented in the MR-MEGA software to investigate the impact of ancestry and the use of a universal blood-based test to define GDM status on heterogeneity in allelic effects on GDM at lead SNVs. We modelled allelic log-ORs across GWAS in a linear regression framework, weighted by the inverse of the variance of the effect estimates, incorporating a covariate indicating whether GDM status was defined by a universal blood-based test (Supplementary Material, Table S1) in addition to the three axes of genetic variation.

Genetic risk score of T2D on GDM

We considered lead SNVs at 237 previously reported loci for T2D from the DIAMANTE Consortium (16) obtained from a multi-ancestry meta-analysis of 180 834 cases and 1 159 055 controls (48.9% non-European ancestry). For each of the 222 SNVs that were reported in our multi-ancestry meta-analysis, we compared association summary statistics (risk allele, other allele, log-OR and *P*-value) for GDM and those reported for T2D. We excluded lead SNVs for T2D that also attained genome-wide significance for GDM. For the remaining SNVs, we regressed the log-ORs for GDM (weighted by their corresponding variances) on the log-OR for T2D, as implemented in `grs.summary` function (57) of the `gtx` package in R version 3.4.2. We estimated the percentage of GDM variance explained, as measured by pseudo R^2 .

Genetic correlation between GDM and glycaemic traits

We used LD Hub (58) to perform LD score regression (59) of the European ancestry association summary statistics for GDM on other glycaemic traits. We included T2D (33), fasting glucose (60), fasting insulin (60), fasting proinsulin (60), glucose 2 h post oral glucose tolerance test (adjusted for BMI) (61), HbA1c (62), HOMA-B (63), HOMA-IR (63), BMI (64), birth weight (41) and alanine aminotransferase (42). European ancestry association summary statistics for GDM were filtered so that only SNVs with minor allele frequency >0.01 was included before performing the LD score regression. Genetic correlations between the different glycaemic traits were obtained from the LD Hub lookup centre. Visualization was performed using the R package `ggplot2` (65) in R version 3.6.1.

Enrichment of GDM association signals in genomic annotations

We mapped each SNV across the genome to three categories of functional and regulatory annotations. First, we considered genic regions, as defined by the GENCODE Project (34), including protein-coding exons, and 3' and 5' UTRs as different annotations. Second, we considered chromatin immuno-precipitation sequence (ChIP-seq) binding sites for 165 transcription factors: 161 proteins from the ENCODE Project (35) and four additional factors assayed in primary pancreatic islets (36). Third, we considered 13 unique and recurrent chromatin states, including promoter, enhancer, transcribed, and repressed regions, in four diabetes-relevant tissues (37): pancreatic islets, liver, adipose and skeletal muscle. This resulted in a total of 220 genomic annotations for enrichment analyses.

We tested for genome-wide enrichment of GDM associations that map to genomic annotations using `fGWAS` (66). To do this, we approximated the Bayes' factor in favour of GDM association for the *j*th SNV by

$$A_j = \exp \left[\frac{D_j - 4 \ln K_j}{2} \right],$$

where D_j is the deviance across K_j contributing GWAS contributing to the multi-ancestry meta-regression (27). We first considered each annotation separately and identified those with significant enrichment ($P < 0.05$). We then used an iterative approach to identify a joint model of enriched annotations from this set. At each iteration, we dropped the annotation from the joint model that minimized the reduction in the penalized likelihood. We continued until no additional annotations worsened the fit of the joint model at nominal significance ($P < 0.05$). We next used the cross-validation likelihood because the significance of parameter estimates from the penalized likelihood cannot be assessed using standard statistical approaches. For the selected joint model, we identified the penalty that maximized the cross-validation likelihood. Finally, we dropped any annotations from the joint

model that resulted in a decrease in the cross-validation likelihood.

Annotation informed fine-mapping of the *HKDC1* locus

At the *HKDC1* locus, we calculated the posterior probability of driving the GDM association for each SNV under an annotation-informed prior model derived from the globally enriched functional and regulatory annotations identified by fGWAS. Specifically, for the j th SNV at the locus, the posterior probability $\pi_j \propto \gamma_j A_j$, where A_j is the Bayes' factor in favour of GDM association from the meta-regression, derived above. In this expression, the relative annotation-informed prior for the SNV is given by

$$\gamma_j = \exp \left[\sum_k \hat{\beta}_k z_{jk} \right],$$

where the summation is over the enriched annotations, $\hat{\beta}_k$ is the estimated log-fold enrichment of the k th annotation from the final joint model and z_{jk} is an indicator variable taking the value 1 if the j th SNV maps to the k th annotation, and 0 otherwise. We derived a 99% credible set (67) for the locus by: (i) ranking all SNVs according to their posterior probability π_j ; and (ii) including ranked SNVs until their cumulative posterior probability attained or exceeded 0.99.

We conducted a look-up of 99% credible set variants at the *HKDC1* locus for significant ($q < 0.05$) cis-expression quantitative trait loci (eQTLs) across tissues in the GTEx Project (44). We reported only those 99% credible variants that were the lead SNV for the eQTL signal.

MR assessment of the effects of metabolic traits on GDM risk

We systematically searched the MR-Base GWAS catalogue (<https://www.mrbase.org>) for metabolic measures. We selected all subcategories of metabolites, which included 'amino acid', 'carbohydrate', 'cofactors and vitamins', 'energy', 'fatty acid', 'keto acid', 'lipid', 'metabolite salt', 'metabolites ratio', 'NA', 'nucleotide', 'peptide', 'protein', 'unknown metabolite' and 'xenobiotics'. We also selected the following subcategories of risk factors: 'anthropometric', 'hormone', 'immune system', 'kidney' and 'metal'. We identified European ancestry GWAS in MR-Base for each selected metabolic trait. Where more than one GWAS was available for a trait, we gave preference to women-specific studies with the largest sample sizes and numbers of SNVs. Any GWAS undertaken only in men were excluded.

For each metabolic trait with more than five genetic instruments, we conducted MR analyses using a 'mixture of experts' (MoE) machine learning approach (68). This approach maximizes statistical power whilst minimizing the impact of horizontal pleiotropy by combining four instrument selection approaches to 14 different MR models. The four approaches for selecting genetic instruments using MoE were: (i) 'top hits' corresponding to independent variants associated at genome-wide

significance ($P < 5 \times 10^{-8}$, $r^2 < 0.001$ using 1000G CEU as the reference population); (ii) 'directional filtration' that removed instruments from 'top hits' that are likely to be related to the outcome through reverse causation using Steiger filtering (69); (iii) 'heterogeneity filtering' that removed instruments from 'top hits' that make a substantial contribution to Cochran's Q statistic ($P < 0.05$); and (iv) combined 'directional filtration' and 'heterogeneity filtering'. The 14 MR models were: seven mean-based methods (inverse variance weighting with fixed effects, IVW random effects, MR-Egger fixed effects, MR-Egger random effects and the three Rucker estimates), three median-based methods (simple, weighted and penalized median estimator) and four mode-based methods (simple and weighted mode, each weighted with or without the assumption of no measurement error in the exposure estimates). The best combination of instrument selection-MR method was identified using a variable predicted by MoE, scaled between 0 and 1, where 1 indicates best performance.

For metabolic traits with five or fewer genetic instruments, the MoE approach could not be applied because many of the MR models require larger numbers of SNVs. For these metabolic traits, we used either the Wald ratio estimate (one SNV) or the inverse-variance weighted estimate (between two and five SNVs).

All analyses were conducted in R version 3.6 using the packages 'TwoSampleMR' (version 0.5.4) and 'MRInstruments' for the MR analyses and 'EpiCircos' (<https://github.com/mattlee821/EpiCircos>).

Supplementary Material

Supplementary Material is available at HMGJ online.

Data Availability

Meta-analysis summary statistics can be downloaded from: https://tools.gi.ut.ee/tools/GENDIP_PervjakovaEtAl2022.txt.gz.

Acknowledgements

ALSPAC. Core funding for ALSPAC is provided by the UK Medical Research Council and Wellcome (217065/Z/19/) and the University of Bristol. Genotyping of the ALSPAC maternal samples was funded by Wellcome (WT088806) and the offspring samples were genotyped by Sample Logistics and Genotyping Facilities at the Wellcome Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe. A comprehensive list of grants funding is available on the ALSPAC website (<http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf>). We are extremely grateful to all the families who took part in ALSPAC, the midwives for their help in recruiting them and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research

scientists, volunteers, managers, receptionists and nurses.

AM. We thank the participants of the ANDIS-MDC study. The study is supported by Swedish Research Council (project grant 521-2010-3490 and infrastructure grants 2010-5983, 2012-5538 and 2014-6395). The study is supported by Crafoord foundation (project grant 20200891) and Hjärt-Lungfonden (project grant: 20180522).

BIB. Born in Bradford (BiB) data used in this research was funded by Wellcome (WT101597MA), a joint grant from the UK Medical Research Council (MRC) and UK Economic and Social Science Research Council (ESRC) (MR/N024397/1), the British Heart Foundation (CS/16/4/32482) and the National Institute for Health Research (NIHR) under its Collaboration for Applied Health Research and Care (CLAHRC) for Yorkshire and Humber and the Clinical Research Network (CRN). BiB is only possible because of the enthusiasm and commitment of the Children and Parents in BiB. We are grateful to all the participants, teachers, school staff, health professionals and researchers who have made Born in Bradford happen.

BOTNIA. The Botnia Study (L.G., T.T.) have been financially supported by grants from Folkhälsan Research Foundation, the Sigrid Juselius Foundation, The Academy of Finland (grant nos 263401, 267882, 312063, 336822 to L.G.; 312072 and 336826 to T.T.), University of Helsinki, Nordic Center of Excellence in Disease Genetics, EU (EXGENESIS), MOSAIC FP7-600914, Ollqvist Foundation, Swedish Cultural Foundation in Finland, Finnish Diabetes Research Foundation, Foundation for Life and Health in Finland, Signe and Ane Gyllenberg Foundation, Finnish Medical Society, Paavo Nurmi Foundation, State Research Funding via the Helsinki University Hospital, Perklén Foundation, Närpes Health Care Foundation and Ahokas Foundation. The study has also been supported by the Ministry of Education in Finland, Municipal Health Care Center and Hospital in Jakobstad and Health Care Centers in Vasa, Närpes and Korsholm. The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement (n° 269045). The study is also supported by Crafoord foundation (project grant 20200891) and Hjärt-Lungfonden (project grant: 20180522) (R.B.P.). The skilful assistance of the Botnia Study Group is gratefully acknowledged.

DECODE. We thank the women who have participated in the deCODE study.

EstBB. This study was funded by the European Union through the European Regional Development Fund (Project No. 2014-2020.4.01.15-0012 and Project No. 2014-2020.4.01.16-0125); by the European Union through Horizon 2020 research and innovation programme under grant agreement (No. 101017802 and 810645); and by the Estonian Research Council grants PUT (PRG687, PRG1291). Data analyses were carried out in part in

the High-Performance Computing Center of University of Tartu. We thank participants and support staff of Estonian Biobank.

FINNGEDI. The EPx-GDM study has been supported by the French government, managed by the National Research Agency (ANR) under the Future Investment Program (PIA) with the references ANR-10-LABX-0046 (European Genomic Institute for Diabetes [E.G.I.D]) and ANR-10-EQPX-07-01 (LIGAN-PM). This work was also supported by the ANR-16-CE17-0017-01, by the Fondation Francophone pour la Recherche sur le Diabète (FFRD) that is sponsored by the Fédération Française des Diabétiques (FFD), Abbott, AstraZeneca, Eli Lilly, Merck Sharp & Dohme (MSD) and Novo Nordisk, and by the National Center for Precision Diabetic Medicine—PreciDIAB, which is jointly supported by the French National Agency for Research (ANR-18-IBHU-0001), by the European Union (FEDER), by the Hauts-de-France Regional Council and by the European Metropolis of Lille (MEL). The FinnGeDi study was supported by: Academy of Finland, Diabetes Research Foundation, Foundation for Pediatric Research, Juho Vainio Foundation, Novo Nordisk Foundation, Signe and Ane Gyllenberg Foundation, Sigrid Jusélius Foundation, Yrjö Jahnsson Foundation for Pediatric Research, Finnish Diabetes Medical Foundation, Research Funds of Oulu University Hospital (state grants), Research Foundation, Novo Nordisk Foundation, funds of Helsinki University Hospital (state grants), Medical Research Center Oulu and National Institute for Health and Welfare (Finland).

ERF. We are grateful to all participants and their relatives, to general practitioners and neurologists for their contributions, to P. Veraart for her help in genealogy, to Jeannette Vergeer for the supervision of the laboratory work and to P. Snijders for his help in data collection. The ERF study was supported by grants from the Netherlands Organisation for Scientific Research (Pionier, 047.016.009, 047.017.043), Erasmus MC and the Centre for Medical Systems Biology (CMSB; National Genomics Initiative). Ayse Demirkan was supported by NWO (VENI-91616165), WCRF-2017/1641 and H2020-SC1-2019-874739.

ESTGDM. This study was supported by the Estonian Research Council grants (IUT20-43 and PRG712). We are grateful to Dr Anne Kirss and midwife Mrs Laura Lauren from Women's Clinic, Tartu University Hospital, for interviewing the participants and collecting the blood samples.

GEN3G. Gen3G has been supported by an American Diabetes Association accelerator award #1-15-ACE-26 (M.F.H.); Fonds de Recherche du Québec en Santé #20697 (M.F.H.); Canadian Institute of Health Research #MOP 115071 (M.F.H.) and Diabète Québec grants (P.P. and L.B.). Gen3G research team thanks all Gen3G participants.

GIFTS. This study was supported by Medical Research Council [Clinical Research Training Fellowship (G0800441)], the European Union (FP7 EU grant: 83599025) and the Diabetes Association of Bangladesh. The authors are

indebted to all the participants from Dhaka (Bangladesh) and London (UK).

HONGKONG. The work for CUHK GDM cohort was supported by the Research Grants Council General Research Fund (CUHK 473408, CUHK 471713, 14118718). We also acknowledge support from the RGC Theme-based Research Scheme (T12-402/13 N) and Research Impact Fund (R4012-18), the Chinese University of Hong Kong-Shanghai Jiao Tong University Joint Research Fund and the Croucher Foundation Senior Medical Research Fellowship.

NFBC Cohort 1966 31y follow-up study. We thank all cohort members and researchers who participated in the 31 years study. We also wish acknowledge the work of the NFBC project centre. NFBC1966 received financial support from University of Oulu (Grant nos 65354 and 24000692); Oulu University Hospital (Grant nos 2/97, 8/97 and 24301140); Ministry of Health and Social Affairs (Grant nos 23/251/97, 160/97, 190/97). National Institute for Health and Welfare, Helsinki (Grant no. 54121); Regional Institute of Occupational Health, Oulu, Finland (Grant nos 50621, 54231, 54121); Regional Institute of Occupational Health, Oulu, Finland (Grant no. 50621, 54231); and ERDF European Regional Development Fund (Grant no. 539/2010 A31592). The NFBC team is acknowledging the following Funding/Support: This work was supported by the European Union's Horizon 2020 research and innovation programme under grant agreement (No. 633595) (DynaHEALTH) and grant agreement No.733206 (LifeCycle) H2020-824989 EUCANCONNECT, H2020-873749 LongITools, H2020-848158 EarlyCause and the JPI HDHL, PREcisE project MR/S03658X/1, ZonMw the Netherlands (no. P75416); the academy of Finland EGEA-project (285547).

NFBC Cohort 1966 46y follow-up study. We thank all cohort members and researchers who participated in the 46 years study. We also wish acknowledge the work of the NFBC project centre. NFBC1966 received financial support from University of Oulu (Grant no. 24000692); Oulu University Hospital (Grant no. 24301140); ERDF European Regional Development Fund (Grant no. 539/2010 A31592).

NFBC Cohort 1986. We thank all cohort members and researchers who have participated in the study. We also wish acknowledge the work of the NFBC project centre. NFBC1986 received financial support: EU QLG1-CT-2000-01643 (EUROBLCS) (Grant no. E51560); NorFA (Grant no. 731, 20056, 30167); USA/NIH 2000 G DF682 (Grant no. 50945). The NFBC team acknowledges the following funding/support: the European Union's Horizon 2020 research and innovation programme under grant agreement (No. 633595) (DynaHEALTH) and grant agreement (No.733206) (LifeCycle) H2020-824989 EUCANCONNECT, H2020-873749 LongITools, H2020-848158 EarlyCause and the JPI HDHL, PREcisE project, ZonMw the Netherlands (no. P75416); the academy of Finland EGEA-project (285547).

PREDO. The PREDO study was supported by Academy of Finland, EVO research funding (A special Finnish

state subsidy for health science research), Finnish Medical Foundation, Jane and Aatos Erkko Foundation, Päivikki and Sakari Sohlberg Foundation, University of Helsinki Research Funding, Jalmari ja Rauha Ahokas Foundation, Yrjö Jahnsson Foundation and Juho Vainio Foundation.

SNUH. This work was supported by the Korea Health 21 R&D Project, Korean Ministry of Health and Welfare (grant no. 00-PJ3-PG6-GN07-001). This study was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute, funded by the Ministry of Health & Welfare, Republic of Korea to S.H.K. (grant number HI15C3131).

STORK. The STORK study received additional funding from the Norwegian Diabetes Association, the Norwegian Odd Fellow Research Fund and Johan Selmer Kvanes' Endowment for Research in Diabetes.

STORKG. Acknowledge Hormone Laboratory, Oslo University Hospital for DNA-extraction.

UKBB. UK Biobank analyses were conducted using the UK Biobank resource under application 11867.

VIVA. Grants from the US National Institutes of Health (R01 HD034568, UH3 OD023286). Project Viva is thankful to all Project Viva participants for their participation to research over many years.

Individual author acknowledgements. D.A.L. is supported by the European Research Council (669545); US National Institute for Health (R01 DK10324) and British Heart Foundation (CH/F/20/90003). D.A.L. and M.C.B. work in a Unit supported by the University of Bristol and the UK Medical Research Council (MC_UU_00011/6) and are supported by a British Heart Foundation Accelerator Award (AA/18/7/34219). R.M.F. and R.N.B. were supported by Sir Henry Dale Fellowship (Wellcome and Royal Society grant: WT104150). R.M.F. is funded by a Wellcome Senior Research Fellowship (WT220390). G.H.M. has received funding from the South-Eastern Health Authority of Norway, the Norwegian Diabetes Association and Nils Normans Minnegave. G.H.M. is supported by the Norwegian Research Council (Post doctoral mobility research grant 287198).

M.C.B. was funded by a Medical Research Council (MRC) Skills Development Fellowship [MR/P014054/1] and a University of Bristol Vice-Chancellor's Fellowship.

The authors acknowledge the use of the University of Exeter High-Performance Computing (HPC) facility in carrying out this work.

Funders and others acknowledged here had no influence on the study design, data collection or interpretation of results. Views expressed are those of the authors and not necessarily any funder or other acknowledged here.

This research was funded in part by Wellcome grants (WT104150 and WT220390). A CC BY or equivalent licence is applied to the author accepted manuscript arising from this submission, in accordance with the grant's open access conditions.

Conflicts of Interest statement. D.A.L. has received support from Roche Diagnostics and Medtronic Ltd for work unrelated to that presented here. M.I.M. has served on advisory panels for Pfizer, NovoNordisk and Zoe Global, has received honoraria from Merck, Pfizer, Novo Nordisk and Eli Lilly, and research funding from Abbvie, Astra Zeneca, Boehringer Ingelheim, Eli Lilly, Janssen, Merck, NovoNordisk, Pfizer, Roche, Sanofi Aventis, Servier and Takeda. M.I.M. is now an employee of Genentech and a holder of Roche stock. G.T., V.S. and K.S. are employees of deCODE genetics/Amgen, Inc.

Authors' Contributions

Central analysis group: N.P., G.-H.M., M.-C.B., T.F., J.P.C., R.M.F., D.A.L., A.P.M., R.M. **Cohort analysis group:** G.M., M.B., T.F., C.A., R.N.B., M.C., G.H., A. Heiskala, A. Joensuu, V.K., S.K., F.T.J.L., J. Liu, S.R.-S., C.H.T., W.T., G.T., S.H., R.B.P. **Cohort sample collection, phenotyping, genotyping and additional analysis:** T.A., J.A., B.B., A.B., F.D., A.D., P.F., K.H., H.H., S.H., A. Hussain, E. Kajantie, E. Keikkala, A.K., J. Lahti, T.L., S.M., C.S., A.T., E.T., R.U., M.V., P.M.V., R.B.P., R.M.F. **Cohort PI:** K.I.B., L.B., C.M.D., S.F., L.G., E.H., G.M.H., G.A.H., H.C.J., M.J., A. Jenum, H.L., R.C.M., O.M., E.O., K.P., P.P., R.B.P., E.Q., S.S., K.S., V.S., T.T., M.-F.H., P.W.F., M.I.M., C.M.L., D.A.L., A.P.M., R.M. **Writing group:** G.-H.M., M.-C.B., C.M.L., R.M.F., D.A.L., A.P.M., R.M.

References

- Metzger, B.E. (2010) Hyperglycaemia and adverse pregnancy outcome (HAPO) study: associations with maternal body mass index. *BJOG An Int. J. Obstet. Gynaecol.*, **117**, 575–584.
- Farrar, D., Fairley, L., Santorelli, G., Tuffnell, D., Sheldon, T.A., Wright, J., van Overveld, L. and Lawlor, D.A. (2015) Association between hyperglycaemia and adverse perinatal outcomes in south Asian and white British women: analysis of data from the born in Bradford cohort. *Lancet Diabetes Endocrinol.*, **3**, 795–804.
- Farrar, D., Simmonds, M., Bryant, M., Sheldon, T.A., Tuffnell, D., Golder, S., Dunne, F. and Lawlor, D.A. (2016) Hyperglycaemia and risk of adverse perinatal outcomes: systematic review and meta-analysis. *Hyperglycaemia and risk of adverse perinatal outcomes: systematic review and meta-analysis. BMJ*, 354.
- Stacey, T., Tennant, P.W.G., McCowan, L.M.E., Mitchell, E.A., Budd, J., Li, M., Thompson, J.M.D., Martin, B., Roberts, D. and Heazell, A.E.P. (2019) Gestational diabetes and the risk of late stillbirth: a case-control study from England, UK. *BJOG An Int. J. Obstet. Gynaecol.*, **126**, 973–982.
- Kwak, S.H., Kim, H.S., Choi, S.H., Lim, S., Cho, Y.M., Park, K.S., Jang, H.C., Kim, M.Y., Cho, N.H. and Metzger, B.E. (2008) Subsequent pregnancy after gestational diabetes mellitus. *Diabetes Care*, **31**, 1867–1871.
- Kramer, C.K., Campbell, S. and Retnakaran, R. (2019) Gestational diabetes and the risk of cardiovascular disease in women: a systematic review and meta-analysis. *Gestational diabetes and the risk of cardiovascular disease in women: a systematic review and meta-analysis. Diabetologia*, **62**, 905–914.
- Vounzoulaki, E., Khunti, K., Abner, S.C., Tan, B.K., Davies, M.J. and Gillies, C.L. (2020) Progression to type 2 diabetes in women with a known history of gestational diabetes: systematic review and meta-analysis. *BMJ*, **369**.
- Farrar, D. (2016) Hyperglycemia in pregnancy: prevalence, impact, and management challenges. *Int. J. Women's Health*, **8**, 519–527.
- American Diabetes Association (2004) Gestational diabetes mellitus. *Gestational Diabetes Mellitus. Diabetes Care*, **2004**(27), s88–s90.
- Zhang, C. and Ning, Y. (2011) Effect of dietary and lifestyle factors on the risk of gestational diabetes: review of epidemiologic evidence. *Am. J. Clin. Nutr.*, **94**(6 Suppl):1975S–1979S.
- Jang, H.C., Min, H.K., Lee, H.K., Cho, N.H. and Metzger, B.E. (1998) Short stature in Korean women: a contribution to the multifactorial predisposition to gestational diabetes mellitus. *Diabetologia*, **41**, 778–783.
- Sattar, N. and Greer, I.A. (2002) Pregnancy complications and maternal cardiovascular risk: opportunities for intervention and screening? Pregnancy complications and maternal cardiovascular risk: opportunities for intervention and screening? *Br. Med. J.*, **325**, 157–160.
- Rich-Edwards, J.W., Fraser, A., Lawlor, D.A. and Catov, J.M. (2014) Pregnancy characteristics and women's future cardiovascular health: an underused opportunity to improve women's health? *Epidemiol. Rev.*, **36**, 57–70.
- Vujkovic, M., Keaton, J.M., Lynch, J.A., Miller, D.R., Zhou, J., Tcheandjieu, C., Huffman, J.E., Assimes, T.L., Lorenz, K., Zhu, X. et al. (2020) Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. *Nat. Genet.*, **52**, 680–691.
- Spracklen, C.N., Horikoshi, M., Kim, Y.J., Lin, K., Bragg, F., Moon, S., Suzuki, K., Tam, C.H.T., Tabara, Y., Kwak, S.H. et al. (2020) Identification of type 2 diabetes loci in 433,540 East Asian individuals. *Nature*, **582**, 240–245.
- Mahajan, A., Spracklen, C.N., Zhang, W., Ng, M.C., Petty, L.E., Kitajima, H., Yu, G.Z., Rueger, S., Speidel, L., Kim, Y.J. et al. (2020) Trans-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. *medRxiv*.
- Mahajan, A., Taliun, D., Thurner, M., Robertson, N.R., Torres, J.M., Rayner, N.W., Payne, A.J., Steinthorsdottir, V., Scott, R.A., Grarup, N. et al. (2018) Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat. Genet.*, **50**, 1505–1513.
- Watanabe, R.M. (2011) Inherited destiny? Genetics and gestational diabetes mellitus. *Genome Med*, **3**(3), 18.
- Powe, C.E. and Kwak, S.H. (2020) Genetic studies of gestational diabetes and glucose metabolism in pregnancy. *Genetic studies of gestational diabetes and glucose metabolism in pregnancy. Curr. Diab. Rep.*, 20.
- Zhang, C., Bao, W., Rong, Y., Yang, H., Bowers, K., Yeung, E. and Kiely, M. (2013) Genetic variants and the risk of gestational diabetes mellitus: a systematic review. *Hum. Reprod. Update*, **19**, 376–390.
- Kawai, V.K., Levinson, R.T., Adefurin, A., Kurnik, D., Collier, S.P., Conway, D. and Stein, C.M. (2017) A genetic risk score that includes common type 2 diabetes risk variants is associated with gestational diabetes. *Clin. Endocrinol.*, **87**, 149–155.
- Powe, C.E., Nodzenski, M., Talbot, O., Allard, C., Briggs, C., Leya, M.V., Perron, P., Bouchard, L., Florez, J.C., Scholtens, D.M. et al. (2018) Genetic determinants of glycemic traits and the risk of gestational diabetes mellitus. *Diabetes*, **67**, 2703–2709.
- Kwak, S.H., Kim, S.H., Cho, Y.M., Go, M.J., Cho, Y.S., Choi, S.H., Moon, M.K., Jung, H.S., Shin, H.D., Kang, H.M. et al. (2012) A genome-wide association study of gestational diabetes mellitus in Korean women. *Diabetes*, **61**, 531–541.
- Auton, A., Abecasis, G.R., Altshuler, D.M., Durbin, R.M., Bentley, D.R., Chakravarti, A., Clark, A.G., Donnelly, P., Eichler, E.E., Flicek,

- P. et al. (2015) A global reference for human genetic variation. A global reference for human genetic variation. *Nature*, **526**, 68–74.
25. McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang, H.M., Fuchsberger, C., Danecek, P., Sharp, K. et al. (2016) A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.*, **48**, 1279–1283.
 26. Gurdasani, D., Barroso, I., Zeggini, E. and Sandhu, M.S. (2019) Genomics of disease risk in globally diverse populations. *Genomics of disease risk in globally diverse populations. Nat. Rev. Genet.*, **20**, 520–535.
 27. Mägi, R., Horikoshi, M., Sofer, T., Mahajan, A., Kitajima, H., Franceschini, N., McCarthy, M.I. and COGENT-Kidney Consortium, T2D-GENES Consortium and Morris, AP (2017) Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Hum. Mol. Genet.*, **26**, 3639–3650.
 28. Scott, R.A., Lagou, V., Welch, R.P., Wheeler, E., Montasser, M.E., Luan, J., Mägi, R., Strawbridge, R.J., Rehnberg, E., Gustafsson, S. et al. (2012) Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat. Genet.*, **44**, 991–1005.
 29. Chen, J., Spracklen, C.N., Marenne, G., Varshney, A., Corbin, L.J., Luan, J., Willems, S.M., Wu, Y., Zhang, X., Horikoshi, M. et al. (2021) The trans-ancestral genomic architecture of glycemic traits. *Nat. Genet.*, **53**, 840–860.
 30. Hayes, M.G., Urbanek, M., Hivert, M.F., Armstrong, L.L., Morrison, J., Guo, C., Lowe, L.P., Scheftner, D.A., Pluzhnikov, A., Levine, D.M. et al. (2013) Identification of HKDC1 and BACE2 as genes influencing glycemic traits during pregnancy through genome-wide association studies. *Diabetes*, **62**, 3282–3291.
 31. Beaumont, R.N., Warrington, N.M., Cavadino, A., Tyrrell, J., Nodzinski, M., Horikoshi, M., Geller, F., Myhre, R., Richmond, R.C., Paternoster, L. et al. (2018) Genome-wide association study of offspring birth weight in 86 577 women identifies five novel loci and highlights maternal genetic effects that are independent of fetal genetics. *Hum. Mol. Genet.*, **27**, 742–756.
 32. Chen, J., Spracklen, C., Marenne, G., Varshney, A., Corbin, L., Luan, J., Willems, S., Wu, Y., Zhang, X., Horikoshi, M. et al. (2020) The trans-ancestral genomic architecture of glycaemic traits. *Astrid Van Hylckama Vlieg*, **14**, 203.
 33. Morris, A.P., Voight, B.F., Teslovich, T.M., Ferreira, T., Segrè, A.V., Steinthorsdottir, V., Strawbridge, R.J., Khan, H., Grallert, H., Mahajan, A. et al. (2012) Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.*, **44**, 981–990.
 34. Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S. et al. (2012) GENCODE: the reference human genome annotation for the ENCODE project. *Genome Res.*, **22**, 1760–1774.
 35. Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Frietze, S., Harrow, J., Kaul, R. et al. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.
 36. Pasquali, L., Gaulton, K.J., Rodríguez-Seguí, S.A., Mularoni, L., Miguel-Escalada, I., Akerman, I., Tena, J.J., Morán, I., Gómez-Marín, C., Van De Bunt, M. et al. (2014) Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat. Genet.*, **46**, 136–143.
 37. Varshney, A., Scott, L.J., Welch, R.P., Erdos, M.R., Chines, P.S., Narisu, N., Albanus, R.D.O., Orchard, P., Wolford, B.N., Kursawe, R. et al. (2017) Genetic regulatory signatures underlying islet gene expression and type 2 diabetes. *Proc. Natl. Acad. Sci. U. S. A.*, **114**, 2301–2306.
 38. Gaulton, K.J., Ferreira, T., Lee, Y., Raimondo, A., Mägi, R., Reschen, M.E., Mahajan, A., Locke, A., Rayner, N.W., Robertson, N. et al. (2015) Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nat. Genet.*, **47**, 1415–1425.
 39. Williams, K., Ingerslev, L.R., Bork-Jensen, J., Wohlwend, M., Hansen, A.N., Small, L., Ribel-Madsen, R., Astrup, A., Pedersen, O., Auwerx, J. et al. (2020) Skeletal muscle enhancer interactions identify genes controlling whole-body metabolism. *Nat. Commun.*, **11**, 2695.
 40. Tyrrell, J.S., Yaghootkar, H., Freathy, R.M., Hattersley, A.T. and Frayling, T.M. (2013) Parental diabetes and birthweight in 236 030 individuals in the UK Biobank study. *Int. J. Epidemiol.*, **42**, 1714–1723.
 41. Warrington, N.M., Beaumont, R.N., Horikoshi, M., Day, F.R., Helgeland, Ø., Laurin, C., Bacelis, J., Peng, S., Hao, K., Feenstra, B. et al. (2019) Maternal and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors. *Nat. Genet.*, **51**, 804–814.
 42. Ward, L.D., Tu, H.C., Quenneville, C.B., Tsour, S., Flynn-Carroll, A.O., Parker, M.M., Deaton, A.M., Haslett, P.A.J., Lotta, L.A., Verweij, N. et al. (2021) GWAS of serum ALT and AST reveals an association of SLC30A10 Thr95Ile with hypermanganesemia symptoms. *Nat. Commun.*, **12**(12), 1–14.
 43. Lee, S.M., Park, J.S., Han, Y.J., Kim, W., Bang, S.H., Kim, B.J., Park, C.W. and Kim, M.Y. (2020) Elevated alanine aminotransferase in early pregnancy and subsequent development of gestational diabetes and preeclampsia. *J. Korean Med. Sci.*, **35**, e198.
 44. GTEx Consortium (2020) The GTEx consortium atlas of genetic regulatory effects across human tissues. *Science*, **369**, 1318–1330.
 45. Guo, C., Ludvik, A.E., Arlotto, M.E., Hayes, M.G., Armstrong, L.L., Scholtens, D.M., Brown, C.D., Newgard, C.B., Becker, T.C., Layden, B.T. et al. (2015) Coordinated regulatory variation associated with gestational hyperglycaemia regulates expression of the novel hexokinase HKDC1. *Nat. Commun.*, **6**, 6069.
 46. Khan, M.W., Priyadarshini, M., Cordoba-Chacon, J., Becker, T.C. and Layden, B.T. (2019) Hepatic hexokinase domain containing 1 (HKDC1) improves whole body glucose tolerance and insulin sensitivity in pregnant mice. *Biochim. Biophys. Acta Mol. Basis Dis.*, **1865**, 678–687.
 47. Hemani, G., Zheng, J., Elsworth, B., Wade, K.H., Haberland, V., Baird, D., Laurin, C., Burgess, S., Bowden, J., Langdon, R. et al. (2018) The MR-base platform supports systematic causal inference across the human phenome. *elife*, **7**, e34408.
 48. Yuan, S. and Larsson, S.C. (2020) An atlas on risk factors for type 2 diabetes: a wide-angled Mendelian randomisation study. *Diabetologia*, **63**, 2359–2371.
 49. Altshuler, D.M., Durbin, R.M., Abecasis, G.R., Bentley, D.R., Chakravarti, A., Clark, A.G., Donnelly, P., Eichler, E.E., Flicek, P., Gabriel, S.B. et al. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature*, **491**, 56–65.
 50. Mitt, M., Kals, M., Pärn, K., Gabriel, S.B., Lander, E.S., Palotie, A., Ripatti, S., Morris, A.P., Metspalu, A., Esko, T. et al. (2017) Improved imputation accuracy of rare and low-frequency variants using population-specific high-coverage WGS-based imputation reference panel. *Eur. J. Hum. Genet.*, **25**, 869–876.
 51. Gudbjartsson, D.F., Helgason, H., Gudjonsson, S.A., Zink, F., Oddsson, A., Gylfason, A., Besenbacher, S., Magnusson, G., Halldorsson, B.V., Hjartarson, E. et al. (2015) Large-scale whole-genome sequencing of the Icelandic population. *Nat. Genet.*, **47**, 435–444.

52. Surakka, I., Sarin, A.-P., Ruotsalainen, S., Durbin, R., Salomaa, V., Daly, M., Palotie, A. and Ripatti, S. (2016) The rate of false polymorphisms introduced when imputing genotypes from global imputation panels. *bioRxiv*, 080770.
53. Cook, J.P., Mahajan, A. and Morris, A.P. (2017) Guidance for the utility of linear models in meta-analysis of genetic association studies of binary phenotypes. *Eur. J. Hum. Genet.*, **25**, 240–245.
54. Devlin, B. and Roeder, K. (1999) Genomic control for association studies. *Biometrics*, **55**, 997–1004.
55. Mägi, R. and Morris, A.P. (2010) GWAMA: software for genome-wide association meta-analysis. *BMC Bioinf.*, **11**, 288.
56. Yang, J., Ferreira, T., Morris, A.P., Medland, S.E., Madden, P.A.F., Heath, A.C., Martin, N.G., Montgomery, G.W., Weedon, M.N., Loos, R.J. et al. (2012) Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.*, **44**, 369–375.
57. Dastani, Z., Hivert, M.F., Timpson, N., Perry, J.R.B., Yuan, X., Scott, R.A., Henneman, P., Heid, I.M., Kizer, J.R., Lyytikäinen, L.P. et al. (2012) Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic traits: a multi-ethnic meta-analysis of 45,891 individuals. *PLoS Genet.*, **8**.
58. Zheng, J., Erzurumluoglu, A.M., Elsworth, B.L., Kemp, J.P., Howe, L., Haycock, P.C., Hemani, G., Tansey, K., Laurin, C., Pourcain, B.S. et al. (2017) LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics*, **33**, 272–279.
59. Bulik-Sullivan, B.K., Loh, P.-R., Finucane, H.K., Ripke, S., Yang, J., Consortium, S.W.G. of the P.G., Patterson, N., Daly, M.J., Price, A.L. and Neale, B.M. (2015) LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.*, **47**, 291–295.
60. Manning, A.K., Hivert, M.F., Scott, R.A., Grimsby, J.L., Bouatia-Naji, N., Chen, H., Rybin, D., Liu, C.T., Bielak, L.F., Prokopenko, I. et al. (2012) A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat. Genet.*, **44**, 659–669.
61. Saxena, R., Hivert, M.F., Langenberg, C., Tanaka, T., Pankow, J.S., Vollenweider, P., Lyssenko, V., Bouatia-Naji, N., Dupuis, J., Jackson, A.U. et al. (2010) Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nat. Genet.*, **42**, 142–148.
62. Soranzo, N., Sanna, S., Wheeler, E., Gieger, C., Radke, D., Dupuis, J., Bouatia-Naji, N., Langenberg, C., Prokopenko, I., Stolerman, E. et al. (2010) Common variants at 10 genomic loci influence hemoglobin A1C levels via glycemic and nonglycemic pathways. *Diabetes*, **59**, 3229–3239.
63. Dupuis, J., Langenberg, C., Prokopenko, I., Saxena, R., Soranzo, N., Jackson, A.U., Wheeler, E., Glazer, N.L., Bouatia-Naji, N., Gloyn, A.L. et al. (2010) New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat. Genet.*, **42**, 105–116.
64. Yengo, L., Sidorenko, J., Kempner, K.E., Zheng, Z., Wood, A.R., Weedon, M.N., Frayling, T.M., Hirschhorn, J., Yang, J. and Visscher, P.M. (2018) Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum. Mol. Genet.*, **27**, 3641–3649.
65. Wickham, H. (2016) *ggplot2: Elegant Graphics for Data Analysis*. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York, Vol. **2016**.
66. Pickrell, J.K. (2014) Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am. J. Hum. Genet.*, **94**, 559–573.
67. Maller, J.B., McVean, G., Byrnes, J., Vukcevic, D., Palin, K., Su, Z., Howson, J.M.M., Auton, A., Myers, S., Morris, A. et al. (2012) Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat. Genet.*, **44**, 1294–1301.
68. Hemani, G., Bowden, J., Haycock, P., Zheng, J., Davis, O., Flach, P., Gaunt, T. and Smith, G.D. (2017) Automating Mendelian randomization through machine learning to construct a putative causal map of the human phenome. *bioRxiv*, 173682.
69. Hemani, G., Tilling, K. and Davey Smith, G. (2017) Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet.*, **13**.